École Nationale Supérieure Louis Lumière
Mémoire de fin d'études
Son 2013

François Heller

Écoute de contrôle au casque en prise de son multicanal

Enceintes virtuelles binaurales
en situation de tournage

Sous la direction de

Matthieu Parmentier
Alan Blum

Rapporteur

Éric Urbain

Saint-Denis, juin 2013

Résumé

Les systèmes de captation de son multicanal (par opposition à stéréophonique ou monophonique) sont actuellement disponibles pour être utilisés en prise de son pour le cinéma. Que ce soit au moment du tournage ou pour les sons additionnels, on peut utiliser des systèmes de captation tri-pistes, quatre pistes, cinq pistes... Afin de récolter de la matière pour alimenter les différents canaux de diffusion de la salle de cinéma. On exploite pour le cinéma des configurations de prise de son assez légères dans lesquelles on ne met pas en place de régie de contrôle, comme dans un car régie pour la retransmission d'un match par exemple. On contrôle ce que l'on enregistre par le biais du casque. Chaque signal peut être placé dans le casque, à gauche et/ou à droite, ou toute autre combinaison linéaire de ces solutions. Il n'y a en revanche, aujourd'hui, aucun moyen mobile de rendu réaliste de la spatialisation telle qu'elle serait restituée dans une salle de projection calibrée selon les normes de diffusion en salle.

Le but de ce travail est de rechercher une solution mobile d'écoute au casque permettant d'apprécier le rendu multicanal au cours de la prise de son. On propose en premier lieu de tester la méthode de convolution de signaux multicanaux par des réponses impulsionnelles « binaurales » préalablement déterminées, correspondant à chaque canal de restitution d'un auditorium de mixage multicanal.

Ce travail met notamment en œuvre des compétences de mesures audio, de traitement du signal et de prise de son. Il est principalement destiné aux preneurs de son et aux intégrateurs/concepteurs de matériel pour la prise de son.

Mots Clés

- Multicanal
- Binaural
- BRIR
- Convolution
- Tournage
- Prise de son

Abstract

Sensor systems used for multichannel sound capture (as opposed to stereo or mono) are currently available for film-making. Whether at the moment of the shooting or for additional sounds, recording devices with up to five tracks can be used in order to supply the different diffusion channels of a cinema. On the field, recording configurations need to be as light as possible, which excludes the use of a control room, contrary to what is done with production trucks for the broadcast of a football match for example. The sound recorded is monitored through headphones. Each signal may be routed to the left and/or right listening channel, or any linear combination of these solutions. However, there is until now no mobile device delivering a realistic spatial rendering as it would be heard in a calibrated movie theater.

The aim of this work is to search for a mobile monitoring solution enabling a preview of multichannel rendering via headphones. The method investigated consists in the convolution of multichannel signals with previously determined "binaural" impulse responses for each diffusion channel of a re-recording stage.

This work involves skills in audio measurement, signal processing and sound engineering. It is therefore primarily intended for sound engineers and integrators/hardware designers of sound equipment.

Key Words

- Multichannel
- Binaural
- BRIR
- Convolution
- Film shooting
- Sound recording

Remerciements

Alan Blum
Matthieu Parmentier
Éric Urbain
Brian Katz
Claude Gazeau
Laurent Millot
Lucie Bourély
Leo Rossi-Roth
Charlie Cabocel
Manuel Deneu
Jean-Pierre Halbwachs
Jean-Marc L'Hôtel
Vincent Magner
François De Morant
Martin Boissau
Michel Bensaïd
Pascal Spitz
Jean Rouchouse
Franck Jouanny
Étienne Hendrickx
Florent Fajole
Philippe Simonet
Greg Faris
Agnès Hominal
Émilien Fabacher
Pascal Heller
Catherine Heller

Table des matières

Résumé **2**

Abstract **3**

Remerciements **4**

Table des matières **5**

Introduction **7**

1 Techniques de prise de son multicanal **9**
1.1 Son multicanal 9
1.2 Systèmes non matricés 9
1.2.1 Croix IRT 9
1.2.2 Double ORTF 10
1.2.3 DPA 5100 10
1.2.4 MMAD 11
1.2.5 Carré Hamasaki 11
1.2.6 Holophone 11
1.2.7 OCT surround 12
1.3 Systèmes matricés 13
1.3.1 Double MS 13
1.3.2 KFM 360 13
1.3.3 Ambisonie 14
1.4 Définition monitoring / écoute de contrôle 15
1.5 Solutions d'écoute disponibles en tournage 16
1.5.1 Aaton 16
1.5.2 Aeta 16
1.5.3 Nagra 16
1.5.4 Sonosax 16
1.5.5 Sound Devices 16
1.5.6 Zaxcom 17
1.6 Pratiques actuelles 17

2 De l'apport de l'outillage binaural **19**
2.1 Audition spatiale 19
2.1.1 Différences de temps 19
2.1.2 Différences d'intensité 20
2.1.3 Indices spectraux 20
2.1.4 HRTF 21
2.2 Techniques binaurales 21
2.2.1 Prise de son binaurale 21
2.2.2 Synthèse binaurale 21
2.3 Limites 22
2.3.1 Individualisation 22
2.3.2 Inversions avant arrière 23
2.4 Enceintes Virtuelles 23
2.4.1 Principe 23
2.4.2 Précisions sur l'opération de convolution 24
2.5 Test préliminaire 26
2.5.1 OH2 26
2.5.2 Description du dispositif technique 27
2.5.3 Compte rendu 27

3 Mesure de BRIR de l'auditorium de mixage cinéma **29**
3.1 Proposition de réalisation pratique 29
3.2 Protocole expérimental 30
3.2.1 Auditorium de l'école 30
3.2.2 Chaine d'acquisition 30
3.2.3 Réglages effectués 32
3.2.4 Stimuli 34
3.3 Obtention des BRIR par déconvolution 38
3.4 Validation des BRIR 39
3.4.1 Méthode 39
3.4.2 Observations 39

4 Exploitation des BRIR **41**
4.1 Captation d'ambiances multicanales 41
4.1.1 Croix IRT 41
4.1.2 Soundfield 43
4.2 Outils logiciels 44
4.2.1 Décodage du format B 44
4.2.2 Convolution 46
4.3 Traitement et résultats 47
4.3.1 Préparation des extraits audio 47
4.3.2 Écoutes comparatives et bilan 48
4.4 Perspectives 49
4.4.1 Portabilité 49
4.4.2 Prise en compte des mouvements de la tête 50
4.4.3 Pistes alternatives 50

Conclusion **52**

Bibliographie **53**
Livres 53
Thèses et mémoires 54
Articles scientifiques 55
Cours et livres blancs 55
Documentation technique 57
Ressources internet 57

Table des illustrations **58**

Annexes **59**
A Détails sur l'expérience OH2 59
B Plan de l'auditorium de l'ENSLL 60
C Norme de diffusion sonore au cinéma 61
D Script Matlab pour la déconvolution 62
E Validation des BRIR : routage du signal 63
F Outils logiciels de convolution 64
F.1 Plug-ins gratuits 64
F.2 Plug-ins payants 65
F.3 Autres applications 65

Introduction

Le cinéma, art de l'illusion, n'a de cesse de proposer un spectacle de plus en plus vraisemblable. Ainsi, la diffusion sonore dans le cadre de la projection cinématographique est en constante évolution depuis l'apparition du son synchrone. Au fil du temps on a pu voir s'agrandir la plage dynamique entre le bruit de fond et les niveaux maximums, augmenter la bande passante fréquentielle et se multiplier le nombre de canaux de diffusion. Ce dernier point a notamment pour objectif de renforcer le sentiment d'immersion dans la scène sonore par l'ajout de canaux « surround ». Le standard le plus répandu de nos jours est le 5.1 avec des canaux centre, gauche, droite, arrière gauche, arrière droit et un canal de renfort de graves. De manière schématique, on utilise majoritairement le 5.1 en cinéma de la manière suivante : la voix est placée en mono au centre, les effets sonores à l'image sont répartis sur le trio gauche, centre, droite, et les ambiances viennent occuper tout l'espace disponible.

En ce qui concerne les ambiances multicanales [1] , le fait de reproduire un espace crédible n'est pas forcément une tâche triviale : une scène sonore est toujours constituée d'une multitude de sources réparties dans l'espace, ne serait-ce que par les phénomènes de réverbérations. Pour concevoir ces ambiances, deux manières de travailler peuvent être envisagées. L'une est de mixer entre eux des éléments mono ou stéréo et de les répartir dans l'espace à l'aide de contrôles dit : « panoramiques », ou en montant les sons sur des pistes dédiées à un canal (par exemple des ambiances destinées aux canaux arrières). L'autre est d'utiliser directement des éléments multicanaux. Bien sûr, un panachage des deux possibilités n'est pas exclu. La première option permet de recréer à sa guise un univers sonore indépendamment de toute contrainte liée à un lieu existant. En revanche, elle nécessite de prendre plus de temps pour trouver ou enregistrer les différentes composantes puis de les mixer harmonieusement entre-elles, là ou la seconde option ne demanderait « que » de trouver « la » bonne prise. Cette dernière n'a été rendue exploitable réellement en diffusion cinéma que récemment, avec l'avènement de la projection numérique et le remplacement de la bobine 35 mm par le DCP (pour Digital Cinema Package) qui n'est autre qu'un ensemble de fichiers informatiques. Le DCP offre la possibilité de lire, à la diffusion, de multiples canaux audio sans compression de données, ni matriçage, ce qui permet de restituer convenablement une prise de son multicanal. La méthode qui consiste à utiliser des prises de son multicanal rend difficile la superposition de deux éléments. En effet, on conçoit aisément que deux

1. On trouve dans la littérature scientifique et technique des exemples où le terme « multicanal » est invariable et d'autres où il s'accorde. Le choix est fait ici de l'accorder.

espaces captés tels quels ne se superposent pas nécessairement de manière heureuse. L'utilisation de prises de son d'ambiances multicanales sur les lieux du tournage n'est pas toujours recommandée lorsque l'on simule un décor, par exemple pour une séquence d'appartement, tournée en studio, sans traitement acoustique, l'espace entendu ne correspondra pas à l'espace vu. On se place évidement ici dans le cas d'une quête de la vraisemblance. Par contre, lorsqu'on tourne dans le véritable décor, par exemple une scène se déroulant dans une boulangerie, tournée dans une véritable boulangerie, une prise de son d'ambiance dans cette même boulangerie, va dans de nombreux cas, apporter une vraisemblance significative. L'espace sonore coïncide parfaitement avec ce qui est vu, le son de la perche « colle » avec le son des ambiances, dans le sens où l'on a les mêmes modifications liées à l'acoustique du lieu. Ce cas de figure se rencontre notamment en tournage pour le documentaire, qui devient le champ d'expérimentation de prédilection de la prise de son multicanal sur le tournage puisqu'on tourne sur les lieux que l'on veut montrer. On est dans la majorité des cas, dans une volonté de montrer le réel, et dans une économie de moyens, qui ne permet pas de disposer d'un temps de post-production son suffisant pour re-fabriquer intégralement des espaces sonores.

Dans ces conditions se pose la question de l'écoute de contrôle de ces contenus multicanaux enregistrés. En configuration traditionnelle de tournage le seul outil disponible est le casque. Il est choisi fermé pour s'affranchir du son ambiant et n'écouter que ce que délivre la section de monitoring de l'appareillage utilisé. Les modèles de référence que l'on peut trouver à la location, sont le Sennheiser HD25 qui est le plus répandu, le Beyerdynamic DT48, « historique », et le Sony MDR-7506, l'« alternative ». Le casque est prévu pour une écoute à deux canaux, or dans le cas étudié on cherche à contrôler entre trois à six canaux. Quelles méthodes mettre en œuvre pour contrôler les prises de son multicanal ? Il parait intéressant de se tourner vers les techniques de restitution binaurales au casque, qui permettent de simuler une distribution d'« enceintes virtuelles ».

1 Techniques de prise de son multicanal

1.1 Son multicanal

On parle de son multicanal lorsque la diffusion de la production sonore est destinée à plus de deux canaux [4], par opposition à la monophonie et à la stéréophonie. Ainsi un système de diffusion multicanal dispose de plus de deux canaux indépendants de diffusion, par exemple : la quadriphonie avec quatre enceintes indépendantes est un système de diffusion multicanal. On retrouve de la même manière des systèmes de prise de son multicanal qui fournissent les différents signaux audio pour alimenter un système de diffusion. Par ailleurs, un mixage multicanal peut très bien être réalisé à partir de sources monophoniques réparties dans l'espace et/ou à partir de sources multicanales, tout comme pour la stéréophonie.

On peut distinguer deux catégories de systèmes de prise de son multicanal : les systèmes non matricés, dont le signal d'un microphone est destiné directement à un canal de diffusion, et ceux matricés, pour lesquels il est nécessaire de recalculer le signal pour les différents canaux à partir des enregistrements effectués. Le signal destiné à chaque canal est une combinaison linéaire des signaux enregistrés.

1.2 Systèmes non matricés

1.2.1 Croix IRT

La croix IRT est généralement constituée de quatre micros cardioïdes dans un plan, orientés à quatre-vingt-dix degrés et espacés de vingt ou vingt-cinq centimètres. On trouve également une version avec des hypercardioïdes qui ne sont espacés que de quatorze centimètres. Comme le double ORTF qui est assez ressemblant, ce système propose une belle sensation d'espace, pendant multicanal de celle d'un couple non coïncident en stéréophonie [3] [66].

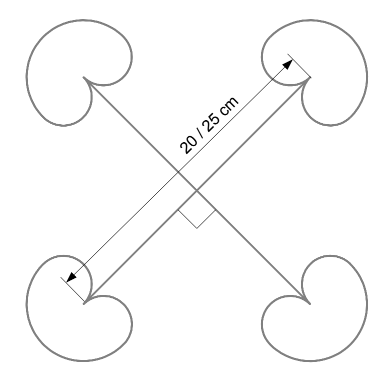


Figure 1 – Croix IRT

1.2.2 Double ORTF

Le double ORTF est constitué de deux couples ORTF dos à dos, espacés d'une trentaine de centimètres. On rappelle que le couple ORTF est constitué de deux capsules cardioïdes espacées de dix-sept centimètres et formant un angle de cent-dix degrés. Schoeps propose également une version appelée ORTF surround avec des capsules hypercardioïdes espacées de dix centimètres entre droite et gauche avec un angle de quatre-vingt degrés et espacées de vingt centimètres entre avant et arrière avec un angle de cent-dix degrés [30].

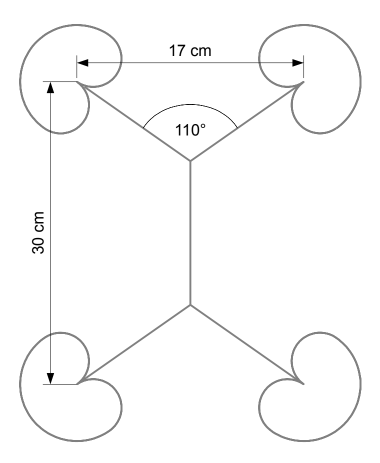


Figure 2 – Double ORTF

1.2.3 DPA 5100

La référence 5100 du constructeur Danois DPA, est un microphone figé constitué de trois capsules coïncidentes directives à l'avant et de

deux capsules omnidirectionnelles espacées pour les arrières. Il est principalement destiné à la captation d'ambiance pour la télévision. Il s'avère particulièrement pratique à mettre en œuvre, étant constitué d'un unique bloc avec un seul connecteur (donc un seul câble), une suspension et une protection au vent intégrée [35].

1.2.4 MMAD

Le terme MMAD est l'acronyme de Multichanel Microphone Array Design. Il s'agit d'une sorte d'algorithme proposé par Michael Williams qui aboutit à un abaque de système en étoile à plusieurs microphones : on choisit le nombre de canaux de diffusion qui doit être égal au nombre de microphones, la directivité voulue de ces derniers, la couverture angulaire voulue pour le triplet central, la configuration préférée pour ce triplet parmi celles qui correspondent, la couverture entre avant et arrière, et enfin, la configuration préférée pour l'arrière entre celles possibles. Il ressort ainsi un abaque correspondant parmi tous ceux possibles [9] [62].

1.2.5 Carré Hamasaki

Le carré Hamasaki est constitué de quatre microphones bi-directionnels placés aux quatre coins d'un carré de deux à trois mètres de côté. Les microphones avant et arrière gauche ont leur lobe de directivité en phase orienté vers la gauche et les microphones avant et arrière droit, vers la droite. L'inconvénient de ce système est qu'il prend beaucoup de place. Son avantage est qu'il permet de capter une ambiance en rejetant la scène frontale, et ainsi limiter la captation de son direct.

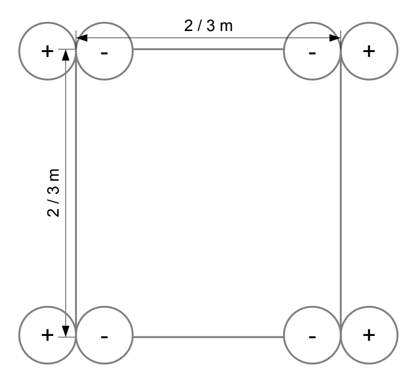


FIGURE 3 – Carré Hamasaki

1.2.6 Holophone

Holophone est une société qui propose différents systèmes de captation multicanal. Tous sont constitués de capsules omnidirectionnelles

encastrées au raz d'un volume ovoïde destiné à reproduire certaines caractéristiques de la tête humaine. En fonction de la version (H2, H2pro, H3D ou H4 super mini) on retrouve une solution 5.1, 7.1 (avec centre, gauche, droit, arrière gauche, arrière droit, centre arrière et élévation), 5.1 plus économique ou 5.1 d'encombrement réduit [56]. Ils présentent l'intérêt d'être facile à mettre en œuvre, d'avoir un rendu en localisation peu précis ce qui facilite le mixage avec d'autres sources [27].



FIGURE 4 – Holophone H2

1.2.7 OCT surround

La version surround de l'OCT, pour Optimised Cardioïd Triangle, propose de placer un microphone cardioïde au centre, deux hypercardioïdes espacés de quarante à quatre-vingt centimètres pointés sur les côtés sur un axe, huit centimètres derrière le centre, pour la restitution de la gauche et de la droite, et deux cardioïdes dirigés vers l'arrière, espacés de vingt centimètres de plus que les microphones latéraux, et placés sur un axe quarante centimètres derrière ces derniers. On lui attribue une grande zone d'écoute ainsi qu'une bonne séparation des secteurs gauche-centre et droite-centre [30].

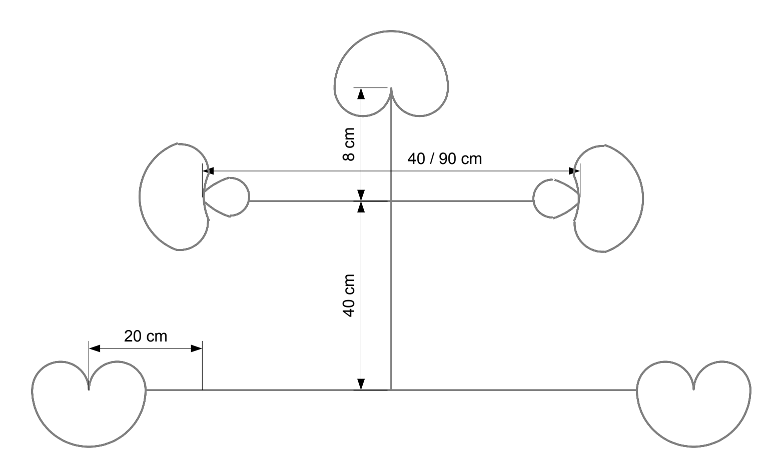


FIGURE 5 – OCT surround

1.3 Systèmes matricés

1.3.1 Double MS

Le double MS (pour Middle Side) est composé de trois capsules coïncidentes : une cardioïde pointée vers l'avant, une autre vers l'arrière et une bi-directionnelle orientée sur les côtés, la partie en phase vers la gauche. Si l'on note le signal des capsules Av, Ar, S, pour avant, arrière et côtés, on retrouve le signal L, C, R, Ls, Rs, (Gauche, Centre, Droit, Arrière gauche, Arrière droit) via les relations :

$$\begin{aligned} L &= Av + S \\ C &= Av \\ R &= Av - S \\ Sl &= Ar + S \\ Sr &= Ar - S \end{aligned}$$

Schoeps propose un boitier hardware ainsi qu'un plug-in software pour faire la conversion Av, Ar, S vers L, C, R, Ls, Rs, LFE. Il est apprécié notamment pour sa qualité de restitution de la localisation, sa souplesse d'utilisation et son faible encombrement.

1.3.2 KFM 360

Schoeps propose également, un dérivé du double MS et de sa sphère KFM 6 [39]. La KFM 360 est une sphère de dix-huit centimètres de diamètre qui constitue un écran acoustique, avec, de manière symétrique à sa gauche et à sa droite, une capsule omnidirectionnelle au raz de la surface. Chacune de ces capsules est surmontée d'un microphone bidirectionnel orienté, en phase, vers l'avant. Il existe également un dématriceur hardware spécifique [30].



Figure 6 – Schoeps KFM 360

1.3.3 Ambisonie

Principe : L'ambisonie est une technique développée par Michael Gerzon. Elle offre la capacité de capter une scène sonore en trois dimensions d'une part, et de la restituer sur un système de diffusion d'autre part, de manière indépendante. Il n'y a pas de lien direct entre système de captation et système de restitution. Elle se base sur la théorie qui décompose le champ sonore en une série d'harmoniques sphériques. L'ambisonie peut se décliner à différents ordres, chacun correspondant à l'ordre des harmoniques sphériques que l'on prend en compte. Plus l'ordre est élevé, plus la précision spatiale est fine. Pour l'ordre un, quatre coefficients d'harmoniques suffisent et l'on trouve des solutions microphoniques sur le marché. L'utilisation des ordres supérieurs, appelés HOA pour High Order Ambisonics, reste encore anecdotique : on peut citer le EigenMike de MHacoustics [2] le SRP de Trinnov Audio [3] et les prototypes de Orange Labs/France Télécom, mais ces solutions ne sont pas compatibles avec une utilisation en sacoche [16].

Format A : Il s'agit du format le plus évident, lorsque l'on parle de captation ambisonique. Il correspond au signal généré par quatre cap-

2. Le EigenMike de MHacoustics est une sphère à trente-deux microphones, intégrant les pré-amplificateurs et convertisseurs, reliée par Ethernet (dans un format propriétaire) à l'interface correspondante (EMIB), vue par un ordinateur sous Windows comme une interface audio firewire [57].

3. Le SRP (Surround Recording Platform) de Trinnov Audio est une structure de fixation avec une géométrie précise, en fer à cheval, pour huit microphones omnidirectionnels, associée à un processeur qui permet de délivrer un signal 5.0 optimisé pour une diffusion suivant la recommandation ITU-R BS.775 [32] [65].

sules cardioïdes disposées en un tétraèdre le plus resserré possible. On trouve dans le commerce des microphones à quatre capsules qui délivrent directement un signal dans ce format, à l'instar de l'Oktava 4-D, du Core Sound TetraMic ou du Soundfield SPS 200 [46].

Format B : Il s'apparente aux harmoniques du premier ordre. On retrouve quatre signaux, W, X, Y et Z qui correspondent à l'information omnidirectionnelle, frontale, latérale, et d'élévation. Le Soundfield ST450 est, par exemple, composé d'un microphone à quatre capsules au format A et d'un boîtier qui délivre les quatre composantes du format B après conversion [47]. On peut passer du format A au format B, avec C_i le signal délivré par la capsule i, par les relations :

$$\begin{aligned} W &= C_1 + C_2 + C_3 + C_4 \\ X &= C_1 + C_2 - C_3 - C_4 \\ Y &= C_1 - C_2 + C_3 - C_4 \\ Z &= C_1 - C_2 - C_3 + C_4 \end{aligned}$$

Il est ensuite possible au mixage, en utilisant soit un processeur hardware dédié soit un plug-in software (tels que Surround zone de Soundfield, Harpex B, Visual Virtual Mic de David Mc Griffy ou l'un des programmes de la suite B2X de Daniel Courville) de revenir au format 5.1 avec quantité de réglages de l'espace possible sous la forme de gestion de microphones virtuels, comme vu dans la partie pratique (cf. section 4.2.1, page 44).

1.4 Définition monitoring / écoute de contrôle

Selon le Dictionnaire encyclopédique du son [4] : « **Monitoring.** De l'anglais monitor qui signifie contrôler, surveiller. Contrôle d'écoute (notion opposée à celle d'écoute récréative). Il s'agit ici d'analyser un signal audio de façon critique, afin d'y repérer d'éventuels défauts d'ordre technique. Ce terme désignait également, sur les magnétophones analogiques l'écoute après bande. » En prise de son multicanal, les critères à contrôler peuvent être [26] :

- le respect de la dynamique
- le respect du timbre
- la profondeur
- la précision de localisation
- la cohérence entre sources réelles et sources virtuelles
- l'angle de prise de son
- l'homogénéité
- le sentiment d'immersion
- le niveau de bruit de fond
- l'absence de distorsion
- l'absence de défauts ponctuels tels que saturation, bruit de vent ou bruit de manipulation

1.5 Solutions d'écoute disponibles en tournage

Tous les systèmes de captation multicanale nécessitent plus de deux pistes d'enregistrement pour être exploités. Il est, en situation mobile, nécessaire d'utiliser un enregistreur multipistes dédié. C'est ce dernier qui va recevoir le signal de chaque canal à enregistrer, et va permettre de le contrôler à l'aide de vumètres, crête-mètres et autres modulo-mètres, en parallèle de l'écoute au casque, gérée grâce à une section de monitoring.

1.5.1 Aaton

Ainsi chez Aaton, qui propose le Cantar X2, il est possible de décider des sources que l'on écoute dans l'oreille gauche, droite ou les deux simultanément dans le menu Out-map avec le grand sélecteur sur 6. Il est décrit dans son manuel comment basculer entre la stéréo avant et arrière d'un double MS et comment écouter la droite et la gauche d'un Format B en jouant sur le mix-down [33].

1.5.2 Aeta

Aeta, avec 4minX intègre des décodeurs MS, double MS, Format A et Format B dans son menu [Monitor] [34].

1.5.3 Nagra

Sur le Nagra VI, chacune des six pistes peut être routée dans le casque à gauche à droite ou au centre via des switch physiques en face avant. On dispose de décodeurs MS pour les paires 1-2, 3-4 et 5-6 et d'un décodeur ambisonique au format A ou B vers stéréo. Enfin, il est possible d'écouter le mix-down stéréo (pistes 7 et 8) dans le menu [audio settings], [monitoring], [monitoring mode] [36] [37].

1.5.4 Sonosax

Sonosax, sur ses enregistreurs SXR4, SX62R et SX-ST : propose une écoute mono, stéréo et un décodage MS grâce au menu [track monitoring], [contextual menu], [monitoring] sur le SXR4, [menu principal] [headphones] sur la section enregistreur du SX62R et à un sélecteur rotatif sur sa section console comme sur les consoles SX-ST [40] [41] [42].

1.5.5 Sound Devices

Sound Devices : En appuyant sur l'encodeur du volume du casque de la 664 on accède au choix du « preset » d'écoute. En activant le switch « Slate/Tone » on accède au menu [HP Preset] qui permet de choisi les sources à gauche, les sources à droite, l'activation d'un décodage MS ou une sommation mono [43].

En travaillant avec la 788, on passe d'un « preset » d'écoute, à un autre, en utilisant l'encodeur rotatif de droite, dans le menu principal. Le menu [Setup Menu], [HP : Monitor Mode], [Edit HP Monitor] permet de créer ses propres « presets ». Il est ainsi permis d'écouter à sa guise les entrées, les pistes et les bus. Il est possible d'utiliser un décodage et une réduction stéréo du format B et MS à une largeur fixée [45].

La 744 présente la même fonctionnalité que sa grande sœur en ajoutant la possibilité d'une écoute « après-bande ». Cette possibilité fait figure d'exception parmi les enregistreurs numériques du marché. Elle présente l'avantage d'être certain d'inscrire des données sur le disque-dur. La contrepartie est que l'écoute est retardée d'environ douze secondes en travaillant à la fréquence d'échantillonage de 48kHz. En effet, il y a dans l'architecture informatique, une mémoire tampon avant l'écriture sur le disque [44].

1.5.6 Zaxcom

Sur les machines Zaxcom Deva 16, Deva 5.8 et Fusion, en suivant l'arborescence [Main Menu], [Setup], [Headphone Options], [Headphone Mix], on peut choisir les sources à écouter sur chaque oreille. Il est également possible de choisir une configuration d'usine. Des inverseurs de phase permettent le décodage MS. De plus, le menu, [Main Menu], [Setup], [Operating Modes], [B-Format] permet d'enclencher un décodeur format B vers stéréo [49] [48].

Sur le plus léger Nomad, on accède à la matrice pour les envois dans le casque en appuyant simultanément sur les encodeurs « MENU » et « HEADPHONES ». Il est ensuite permis de choisir d'écouter à droite et/ou à gauche, les pistes enregistrés, les sorties, les entrées de retours et même l'une des fréquences du réseau d'ordres ZaxNet dont l'émetteur est intégré à l'enregistreur. Il n'y a pas encore de décodeur MS a proprement parler intégré dans le software du Nomad, mais il est possible de bricoler une écoute stéréo à partir d'une source MS via les inversions de phase possible sur les sorties [50].

1.6 Pratiques actuelles

Comme vu précédemment, avec les outils actuellement utilisés en tournage, il est possible de contrôler le bon acheminement de chaque canal sur la piste prévue et la qualité de tous les signaux (défaut de timbre, dynamique, bruit de fond, parasite) en écoutant les pistes séparément en solo. Il est possible également, de contrôler le niveau de modulation de chaque piste, pour toutes les pistes en même temps aux vumètres, de contrôler la présence de bruit de manipulation, de vent, de saturation pour une piste, en cours de prise de son, en écoutant une sommation mono ou stéréo de toutes les pistes. Si l'on n'entend pas de défaut sur l'ensemble des pistes c'est qu'il n'y a de défaut sur aucune des pistes... On a par ailleurs la possibilité d'écouter chaque paire de canaux d'un système multicanal,

en stéréo. On parle de stéréo avant et de stéréo arrière. Une bonne base de départ est de placer le système multicanal à la position dans laquelle serait placé un couple stéréo [27]. Une approche, plus radicale, adoptée notamment par Jean-Marc l'Hôtel, preneur de son, membre actif de l'AFSI (l'Association Française du Son à l'Image), participant au FISM (le Forum International du Son Multicanal), rédacteur de plusieurs papiers sur la prise de son multicanal est de se dire que le multicanal restitue de manière suffisamment fidèle la spatialisation, pour que l'endroit idéal où placer le système de captation soit celui qui corresponde à l'endroit où l'écoute directe, par les oreilles du preneur de son, est optimale. Cela implique d'écouter la scène sonore sans casque pour définir le placement adéquat et de faire « confiance » au matériel.

On remarque dès lors l'impossibilité d'avoir un rendu de la spatialisation telle qu'elle serait restituée dans un auditorium de mixage ou une salle de cinéma.

Après avoir dialogué avec plusieurs chefs opérateurs du son (notamment François de Morant, Vincent Magnier. . .) et inspecté les forums dédiés à la prise de son à l'image [51] [58], on pourrait se risquer à classer les preneurs de son multicanal pour le cinéma en deux catégories : ceux qui ont déjà des habitudes en prise de son multicanal et ceux qui débutent dans ce domaine. Les premiers se sont sans doute déjà habitués aux outils qu'ils ont à disposition et ont dû trouver des repères. Les seconds n'ont pas d'idée de la correspondance entre le son entendu au casque et l'espace restitué dans une salle de projection. Un dispositif permettant d'avoir une approche de rendu de spatialisation devrait donc leur permettre de débuter plus sereinement, sans partir dans l'inconnu. Un tel système devrait de plus permettre à ceux qui ont une habitude avec les outils « traditionnels » d'affiner le placement du système de captation et de passer d'un système de captation à l'autre en gardant les mêmes repères.

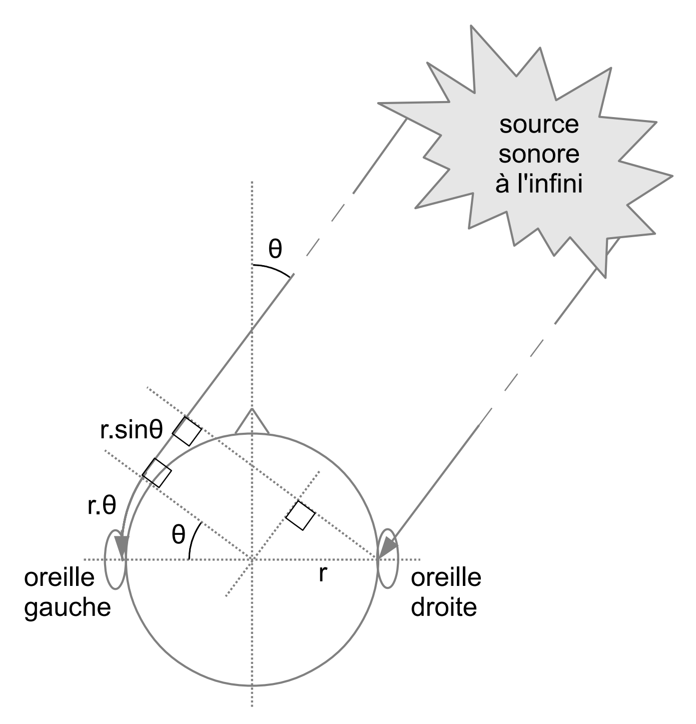
2 De l'apport de l'outillage binaural

2.1 Audition spatiale

À l'écoute d'une source sonore, l'être humain est capable, dans une certaine mesure, de déterminer sa position, sa distance, sa taille, si elle est en mouvement... La localisation d'une source sonore se fait principalement par l'analyse des différences entre ce que perçoivent les deux oreilles. Interviennent notamment le décalage temporel de réception de l'information, ou déphasage, la différence d'amplitude, et les différences spectrales. Il sera décrit ici les mécanismes essentiels liés à l'écoute et aux techniques binaurales. Pour un approfondissement sur l'aspect auditif, voir l'ouvrage [1], qui fait référence en la matière sur les aspects auditif relié à la perception de l'espace.

2.1.1 Différences de temps

On parle de différences interaurales de temps, ou ITD pour « Interaural Time Differences » en anglais. Le front d'onde d'une source placée en dehors du plan médian (plan perpendiculaire à la droite passant par les deux oreilles et passant par le milieu du segment formé par les deux oreilles) atteindra les deux oreilles d'un auditeur avec une petite différence de temps, dépendant de l'angle d'incidence de la source. Le délai maximum du temps d'arrivée entre les deux oreilles correspond à une source entièrement latéralisée à gauche ou à droite. Il avoisine les 0,65 ms et est appelé délai binaural. Pour une fréquence donnée, ce délai peut être interprété comme une différence de phase et renseigner de manière assez précise le système auditif sur la direction d'incidence d'une onde sonore. En revanche, l'information de différence de temps ne permet pas à l'auditeur de faire la distinction entre une source à l'avant et sa source symétrique par rapport à la droite passant par les oreilles, à l'arrière. L'auditeur se sert alors des mouvements de la tête : en effet une source à l'avant et une source à l'arrière vont être perçues comme ayant des directions opposées avec un mouvement de rotation de la tête [21].



Différence de marche entre oreille gauche et droite :

$r.\theta + r.\sin\theta$

FIGURE 7 – Différences interaurales de temps

2.1.2 Différences d'intensité

À partir d'une certaine fréquence, les différences de phase deviennent plus difficiles à exploiter pour le système auditif car elles correspondent à plusieurs périodes de la fréquence en question. C'est alors des différences d'intensité entre les deux oreilles qui prennent le relai (différences interaurales d'intensité ou ILD pour « Interaural Level Differences »). En effet, le volume formé par la tête d'un auditeur forme un obstacle au son, tout particulièrement pour les fréquences les plus élevées. À noter que pour des sources proches de la tête, une différence d'amplitude due à la distance entre les deux oreilles est perceptible (Pour les sources plus éloignées, cette différence est négligeable devant la distance de la source). C'est bien sûr ce mécanisme de localisation des sources sonores qui est largement exploité par les fameux panoramiques d'intensité permettant de placer une source monophonique dans un espace stéréo.

2.1.3 Indices spectraux

La forme particulière du pavillon génère des réflexions, des diffractions et des résonances qui vont modifier le spectre du son arrivant jusqu'au tympan, et ce, en fonction de l'angle d'incidence. Les réflexions sur les épaules et le visage vont elles aussi influencer le spectre sonore en fonction de la position de la source. Ces empreintes spectrales vont per-

mettre notamment de compléter le mécanisme de localisation auditive dans le plan vertical.

2.1.4 HRTF

La combinaison de tous ces effets, constitue, pour une position donnée de source dans l'espace par rapport à l'auditeur, une fonction de transfert. À chaque angle d'incidence correspond donc une fonction de transfert, ou HRTF pour « Head-Related Transfert Function ». Ces filtres sont spécifiques à chaque individu puisqu'ils dépendent de sa morphologie. Actuellement, on s'attache à trouver des points communs entre les HRTF de chacun. Blauert dans les années 70 avait déjà mis en évidence des bandes de fréquence correspondant à des directions privilégiées de l'espace.

2.2 Techniques binaurales

Il y a deux manières de se servir de ces « filtres morphologiques » : soit par la prise de son directe au niveau des oreilles, soit par synthèse en filtrant un son monophonique à l'aide de HRTF gauche et droite préalablement mesurées. Dans les deux cas, on parle alors de techniques binaurales, pour faire référence à l'exploitation de tous ces indices acoustiques captés par nos deux oreilles [2].

2.2.1 Prise de son binaurale

Il suffit de placer un microphone miniature dans chacune des deux oreilles d'une tête pour réaliser une prise de son, dite, binaurale. On peut se servir d'une tête artificielle ou de la tête d'une personne. Dans les deux cas, les microphones enregistreront la scène sonore avec l'empreinte de tous les indices acoustiques relatifs à la tête. Il suffit ensuite d'écouter cet enregistrement au casque, en routant le signal gauche sur le transducteur gauche du casque et celui de droite sur le transducteur droit pour que le cerveau puisse exploiter les indices binauraux. L'auditeur est alors en mesure de percevoir l'espace tridimensionnel de la scène sonore captée, avec par exemple des sources sonores à l'arrière de l'auditeur, au dessus de l'auditeur... On parle d'externalisation. L'utilisation d'une tête artificielle ou d'un mannequin (lorsqu'il s'agit d'une tête artificielle avec un torse) permet de s'affranchir de tous les bruits de déglutition, de respiration et autres. Elle permet aussi de placer la tête à des endroit ou il serait risqué de placer la tête d'un individu. On évite également les mouvements parasites. En revanche l'utilisation de la tête d'un individu permet d'être bien plus discret et réactif. Voir [28] pour une comparaison plus détaillée des différentes méthodes de prise de son binaurale.

2.2.2 Synthèse binaurale

Pour réaliser une synthèse binaurale, on commence par mesurer les HRTF d'un individu, afin d'obtenir une paire de filtres gauche/droite cor-

respondant à chaque direction de l'espace. Il existe différentes méthodes pour cela : dans une chambre anéchoïque, on place des microphones miniatures dans les oreilles de l'individu et on diffuse un stimulus par une enceinte située à une position donnée par rapport à la tête. On répète l'opération un grand nombre de fois pour différents points de l'espace. Les stimuli ainsi enregistrés, on retrouve la réponse impulsionnelle pour une position de source donnée, en déconvoluant l'enregistrement par le stimulus d'origine. Une fois obtenues les HRTF pour un maillage de positions de l'espace, on peut déterminer la transformation subie par un son, pour une direction donnée, par interpolation entre les positions « voisines ». Voir le travail de thèse de Véronique Larcher [19] pour approfondir le sujet d'un point de vue technique. On peut dès lors synthétiser une scène sonore tridimensionnelle en convoluant un élément sonore par la réponse impultionnelle correspondant à la direction choisie. On dispose alors d'une sorte de « panoramique binaural », qui permet de spatialiser artificiellement un signal monophonique dans un espace en trois dimensions. Comme précédemment il est nécessaire que l'écoute soit faite au casque afin que chaque oreille ne reçoive que le signal lui étant destiné.

2.3 Limites

2.3.1 Individualisation

Une première limite des techniques binaurales est due au fait que les HRTF sont propres à chaque individu puisqu'elles dépendent de sa morphologie. Ainsi, en théorie, une prise de son binaurale, ou une scène sonore synthétisée à partir d'un jeu de HRTF ne sera compatible à l'écoute qu'avec la personne qui a prêté sa tête, tête qui présente une morphologie unique et particulière, à la manière des empreintes digitales. L'écoute d'un contenu binaural, avec « les oreilles d'un autre », peut donc être sujette à des distorsions de l'espace généré. Il se trouve cependant que d'un individu à l'autre, ces jeux de filtres présentent des similitudes que l'on tente de dégager, pour trouver des HRTF génériques qui correspondraient au plus grand nombre [18].

On peut contrebalancer cette limitation en mettant en avant la faculté d'un individu à s'habituer relativement rapidement à l'utilisation de HRTF qui ne lui correspondent pas à l'origine. On constate un apprentissage plus rapide encore s'il comprend une phase interactive, par la manipulation de sources sonores virtuelles au cours d'un « jeu » d'exploration de la « sphère auditive » par exemple [12]. On peut également mentionner l'existence de banques de HRTF, constituées notamment sur l'impulsion de l'IRCAM, l'Institut de Recherche et de Coordination Acoustique Musique, dans lesquelles on finit toujours par trouver un jeu qui correspond plus ou moins. Pour éviter la tâche fastidieuse et coûteuse (voire inabordable de par l'utilisation d'une chambre anéchoïque) de la mesure de ses propres fonctions de transfert.

2.3.2 Inversions avant arrière

L'autre principal défaut imputé aux techniques binaurales est la présence d'inversions avant/arrière. En effet il a été vu en 7 (page 20) que c'est en partie grâce aux mouvements de la tête, qu'un auditeur fait la discrimination avant arrière. Or lors d'une écoute au casque, il n'y a pas de mouvement relatif possible entre la tête de l'auditeur et la scène sonore. C'est pourquoi, appuyé par le fait de ne pas voir de source sonore devant soi, le cerveau aura tendance à imaginer la source derrière la tête.

Pour réduire ce défaut, on peut mettre en place un système de « head tracking » qui consiste à mesurer les mouvements de la tête lors de la diffusion binaurale, pour ajuster la spatialisation de la scène en fonction des mouvements de le tête et permettre alors un mouvement relatif de la tête par rapport à la scène. Ce procédé réduit le nombre d'inversions. Il nécessite des systèmes de capteurs d'orientation à placer sur le casque. Différentes technologies existent (champ magnétique, ultrasons, optique), souvent assez couteuse et complexe à mettre en place, même si des formules plus abordables commencent à être envisagées à base de capteurs gyroscopiques par exemple (souris, smartphone…).

2.4 Enceintes Virtuelles

2.4.1 Principe

On a vu que les techniques binaurales permettent, dans une certaine mesure, de créer l'illusion d'une source sonore émettant depuis une direction choisie dans l'espace. En prenant un flux audio et en lui appliquant une HRTF par convolution on obtient un flux audio « binauralisé ». On peut parler de haut-parleur virtuel, ou d'enceinte virtuelle à laquelle on assigne ce flux audio. Cette enceinte virtuelle, simulée via le casque de l'auditeur, délivre alors le flux audio semblant provenir de la position correspondant à celle des HRTF. Pour une paire de HRTF gauche/droite, correspond une position dans l'espace 3D.

Le procédé décrit ici utilise la synthèse binaurale et nécessite d'avoir les HRTF correspondants aux positions voulues pour les sources virtuelles. En général, les mesures de banque de HRTF ne sont faites qu'à une distance fixe de l'auditeur, distance correspondant au dispositif de mesure, et l'on se contente d'un maillage sur les directions. Cela se justifie facilement pour des mesures effectuées en chambre anéchoïque, étant donnée que dans cette condition assimilable au champ libre il ne reste plus que la décroissance du niveau sonore comme indice de distance sonore. Toujours est-il que, lors de la synthèse, pour générer une impression de distance il faudra alors recourir à l'ajout d'une réverbération artificielle, en jouant sur le rapport entre son direct et son réverbéré.

Une autre approche peut consister à enregistrer une enceinte réelle depuis un mannequin de prise de son binaurale. L'écoute au casque d'un tel enregistrement aura lui aussi pour effet de faire entendre le contenu

audio comme venant de la direction correspondant à la position de la source au moment de la captation. Dans ce cas de figure, l'effet de salle est également capté par le système et ainsi la perception de la distance se fait naturellement puisque tous les indices de réverbération sont enregistrés de manière cohérente. En revanche celle-ci est à peu près figée, comme pour n'importe quelle prise de son au couple. On retrouve l'analogie avec la prise de son fractionnée qui est spatialisée avec des panoramiques et des réverbérations artificielles par opposition à la prise de son globale, où les impressions de profondeur et de spatialisation sont fixées dès la captation. Tout preneur de son s'accordera à dire que pour ces dernières, l'espace semble plus naturel et que le timbre y est mieux respecté.

Dans le cadre d'un travail sur un outil de monitoring, on préfèrera par conséquent se tourner vers la solution de simulation d'enceintes virtuelles par la prise de son pour bénéficier de l'acoustique réelle d'un lieu. Toutefois, si cette simulation doit passer pour chaque son à spatialiser par la diffusion dans un auditorium de mixage pour obtenir le contenu binaural, il n'y aurait aucun intérêt, ou en tout cas, aucune application au contexte du tournage. Le principe est donc ici de mettre en place une solution que l'on pourrait qualifier d'hybride : il ne s'agit pas d'enregistrer le contenu audio depuis l'auditorium de mixage, mais de capter l'empreinte de cet auditorium, sous forme de réponse impulsionnelle, à partir d'un système d'enregistrement binaural. Ces empreintes (auxquelles on donne le nom de BRIR pour « Binaural Room Impulse Response »), contiennent à la fois un filtrage de type HRTF et la marque de l'effet de salle. Elles devront être captées pour chaque canal de diffusion. Il sera alors ensuite possible de traiter n'importe quel flux audio en lui donnant l'empreinte d'un canal, comme une enceinte virtuelles encore une fois. Ce procédé sera exposé plus en détail dans la partie 3 de ce mémoire, après quelques précisions sur les outils en jeu, ainsi que la description d'un test d'écoute préliminaire.

2.4.2 Précisions sur l'opération de convolution

Convolution directe

Un filtre quelconque peut être considéré comme un système linéaire invariant dans le temps. La réponse impulsionnelle de ce système est la sortie délivrée lorsqu'on lui applique une impulsion en entrée. Un filtre est entièrement caractérisé par sa réponse impulsionnelle. On a, avec s la sortie, e l'entrée, h la réponse impulsionnelle et $*$ l'opération du produit de convolution :

$$s = h * e$$
$$s(t) = \int_{-\infty}^{+\infty} h(\tau).e(t-\tau).\mathrm{d}\tau$$

En audionumérique, en pratique, la longueur de la réponse impulsionnelle sera généralement finie. Pour une longueur de N échantillons, on peut alors écrire :

$$s[n] = \sum_{k=0}^{N-1} h[k].e[n-k]$$

Ce qui permet d'observer le nombre d'opérations nécessaires au calcul d'un échantillon de sortie $s[n]$ en fonction des échantillons d'entrée et de la longueur N de la réponse impulsionnelle : N multiplications et $(N-1)$ additions, soit $2N-1$ opérations. Cela étant à répéter 48000 fois par seconde pour une fréquence d'échantillonnage de 48kHz. Pour une réponse impulsionnelle d'un quart de seconde (soit 12000 échantillons à 48kHz) cela représente déjà $24000-1$ opérations pour un échantillon, soit plus d'un milliard d'opérations élémentaires par seconde.

Transformée de Fourier rapide

Les calculs de convolution peuvent donc être lourds et fastidieux. La transformée de Fourier définie comme la fonction qui à $a(t)$ associe :

$$A(f) = \int_{-\infty}^{+\infty} a(t).\mathrm{e}^{2\pi jft}.\mathrm{d}t$$

Elle permet de passer, avec j le nombre complexe tel que $j^2 = -1$, d'une représentation temporelle à une représentation fréquentielle des signaux et de transformer le produit de convolution en un simple produit. Soient $E(f)$ la transformée de Fourier de $e(t)$, $S(f)$ celle de $s(t)$ et $H(f)$ celle de $h(t)$. On obtient $S = H.E$ avec H la fonction de transfert du filtre. On peut repasser de la transformée, au signal temporel grâce à la transformée inverse qui à $B(f)$ associe :

$$b(t) = \int_{-\infty}^{+\infty} B(f).\mathrm{e}^{2\pi jft}.\mathrm{d}f$$

En numérique, des algorithmes de transformée de Fourier rapide dits FFT pour, « Fast Fourier Transform », optimisent la transformation et permettent des traitements temps réel. Pour convoluer des signaux temporels en temps réel, on peut donc utiliser la FFT pour basculer dans le domaine fréquentiel, réaliser une simple multiplication et faire une transformée inverse par FFT à nouveau pour revenir dans le domaine temporel. L'analyse du gain, en termes de poids en calculs, pour une machine est assez délicate, mais il est communément admis que l'on passe d'une complexité en $O(N^2)$ à une complexité en $O(N.\log(N))$, ceci devenant très vite intéressant, en général à partir de N supérieur à 100 [10] [5].

Utilisation audionumérique

Dans le cadre des techniques du son, l'utilisation de la convolution est répandue. On trouve notamment des réverbérations à convolution, utilisées dans l'objectif de reproduire ou simuler un effet de salle réelle. On trouve également de nombreux simulateurs d'amplificateurs ou de périphériques analogiques. Tous ces outils de traitement audio sont constitués d'un convolueur associé à une réponse impulsionnelle (RI). Les logiciels, ou plug-ins sont donc presque toujours livrés avec une ou plusieurs RI, correspondant(s) au traitement(s) attendu(s). Il est parfois possible d'y rajouter d'autres RI, éventuellement issues de mesures personnelles. La méthode la plus évidente pour obtenir une RI est de soumettre le système à simuler à une impulsion (coup de pistolet chargé à blanc, explosion de ballon de baudruche) et d'enregistrer sa sortie. Mais on ne génère jamais une impulsion parfaite. Une méthode plus efficace (en termes de robustesse de mesure et de rapport signal à bruit notamment) consiste à enregistrer la réponse, *rep*, à un stimulus connu, *stim*. On retrouve la réponse impulsionnelle, *ri*, par déconvolution (ou convolution inverse) : $ri = rep *^{-1} stim$. Comme on peut l'imaginer à la suite des paragraphes précédents, les demandes en ressources de calculs de ce type de traitement sont assez importantes. En revanche, ils ont pour intérêt de pouvoir offrir des traitements de très bonne qualité, à condition que les RI utilisées soit elles-mêmes de qualité.

2.5 Test préliminaire

2.5.1 OH2

L'association OmniHead, qui s'intéresse à l'utilisation du binaural pour produire et diffuser des contenus a été accueillie dans les locaux de France Télécom R & D à Lannion dans le Finistère pour une session d'expériences autour de prises de son binaurales. L'une d'elles, intitulée « procédé OH2 » (pour OmniHead à deux canaux) consiste à faire des prises de son binaurales d'un contenu diffusé en multicanal. Ici il s'agit d'enregistrer la diffusion d'extraits de DVD du commerce en 5.1 avec une paire de micros miniatures placés dans les oreilles d'un mannequin et d'un des expérimentateurs (cf. annexe A, page 59) [29]. Il a été possible de récupérer ces prises de son. Ainsi, en extrayant le contenu audio des DVD correspondants il a été possible de faire un test préliminaire qui consistait à comparer d'une part une diffusion multicanale du contenu d'un DVD, avec une diffusion au casque de la captation binaurale de celle-ci, et d'autre part, les diffusions au casque de la version binaurale et d'une réduction stéréo du 5.1 original. Ces écoutes on l'intérêt de pouvoir donner un aperçu de ce que pourrait fournir une simulation de 5.1 en binaural à partir d'une prise de son dans un vrai lieu de mixage multicanal, plutôt qu'à partir de traitements de synthèse.

2.5.2 Description du dispositif technique

Les éléments de bande son des DVD sont extraits au format AC3 [4] grâce au « freeware [5] » Mac The Ripper, version 2.6.6 [53] et la conversion au format wave est effectuée par l'utilitaire mAC3dec dans sa version 1.6.1 [54]. Les écoutes comparatives se font dans la salle de montage son numéro cinq de l'école Louis Lumière, dans ses nouveaux locaux à Saint-Denis. L'interface avec Protools 10 est une Digi 002 Rack. Un casque, Sennheiser HD25, est directement relié à sa sortie casque, tandis que les sorties lignes sont connectées à un contrôleur de diffusion, SPL surround monitor controller Model 2489, qui gère l'envoi vers les amplificateurs de puissance, Samson servo 170, des enceintes Dynaudio : trois BM5P à l'avant et deux BM5 à l'arrière.

2.5.3 Compte rendu

Comparaison 5.1/binaural : Lors d'un passage de la diffusion en multicanal sur enceintes à la version binaurale au casque, on remarque :

- l'apparition d'un effet de salle
  (le binaural a dû être enregistré dans une salle moins mate que la salle de montage)
- l'apparition des extrêmes graves
  (il n'y a pas de canal LFE dans la salle de montage)
- le rapprochement des sources
- une spatialisation plus floue
- une spatialisation assez bien respectée
  (phénomène plus marqué avec la version Kemar 3)
- le déplacement de certaines sources vers l'arrière
  (phénomène plus marqué avec la version Kemar 5)
- une perte de précision
- une perte de la qualité du timbre

Comparaison stéréo/binaural : Lors d'un passage de la diffusion de la somme en stéréo au casque à la version binaurale au casque, on remarque :

- une externalisation du son
  (passage intra-cranien vers extra-cranien)
- un démasquage de certaines sources
- un gain en spatialisation
- une perte de précision
  (par l'ajout d'un effet de salle)
- une perte de la qualité du timbre
  (comparable à un effet d'éloignement)
- le fait d'« entendre une salle »

4. Le format AC3 est un format audio multicanal, de compression de données, avec pertes, développé par Dolby.

5. Un freeware, ou gratuiciel, est un logiciel propriétaire distribué gratuitement sans toutefois conférer à l'utilisateur certaines libertés d'usage associées au logiciel libre. Ne doit pas être confondu avec logiciel libre ni open source.

On note que ces effets sont susceptibles d'être retrouvés dans le procédé des enceintes virtuelles. Il semble pertinent d'être particulièrement attentif à l'acoustique du lieu (un auditorium sera sans doute plus mat que la salle mise à disposition pour l'expérience « OH2 »), au réglage du mannequin avec le placement des microphones, et à la gestion du renfort de graves.

3 Mesure de BRIR de l'auditorium de mixage cinéma

3.1 Proposition de réalisation pratique

L'utilisation d'enceintes virtuelles en écoute binaurale pourrait donc permettre de progresser dans l'approche d'un rendu de spatialisation. Toutes les salles de projection (en France tout du moins) doivent, pour avoir l'agrément du CNC, satisfaire à un certain nombre de normes. C'est la CST (Commission Supérieure Technique) qui est en charge de contrôler le respect de ces directives. Ainsi, on peut affirmer que d'une salle à l'autre, la projection doit être plutôt ressemblante sans pour autant pouvoir dire que l'on a la possibilité de profiter d'un spectacle identique. Mais la diffusion du contenu multicanal sur un système de diffusion et son enregistrement binaural n'est pas possible dans le cadre d'une utilisation portable temps réel. Il est ainsi décidé de réaliser des BRIR (Binaural Room Impulse Response) d'un auditorium de mixage calibré, pour avoir la possibilité de simuler, par convolution, les différents canaux de la diffusion cinéma. Les BRIR correspondent à la réponse impulsionnelle d'une salle enregistrées par un système binaural (c'est à dire une tête humaine ou artificielle équipée de microphones au niveau des oreilles) pour intégrer des HRTF dans la réponse implulsionnelle mesurée. Ici, pour simuler les enceintes d'une salle de projection, les stimuli vont être délivrés dans la salle par toute la chaîne prévue habituellement (même enceintes, même position, niveau nominal...)

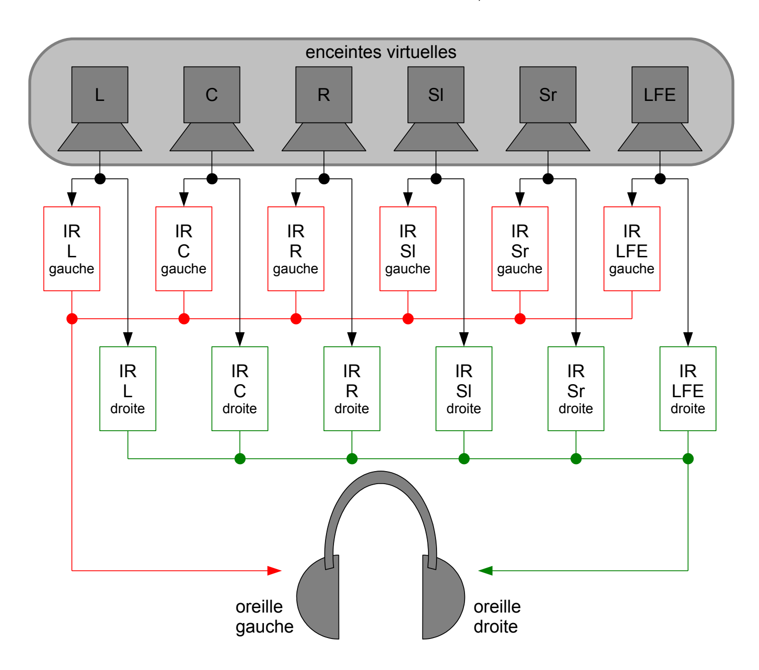


FIGURE 8 – Principe des enceintes virtuelles

3.2 Protocole expérimental

3.2.1 Auditorium de l'école

L'école dispose d'un tout nouvel auditorium de mixage dans ses nouveaux locaux dans la Cité du Cinéma. C'est celui-ci qui a été utilisé pour l'acquisition des réponses impulsionnelles. Il a été installé et réglé par la société 44.1. Il mesure 10,21 mètres de long pour 7,44 mètres de large avec une hauteur sous plafond de 2,70 mètres (cf. figure 26, annexe B, page 60). Il est équipé d'une écoute en 5.1. Il est calibré pour répondre aux normes de diffusion des salles de cinéma décrit dans le recueil de recommandations : [8], avec un niveau sonore de 85dBSPL (pondéré C) par canal, au point d'écoute et une courbe de réponse en fréquence s'inscrivant dans l'abaque ISO X (présenté en annexe C page 61).

3.2.2 Chaine d'acquisition

La chaine d'acquisition est la suivante : un stimulus sous la forme d'un fichier audio est lu par le séquenceur Pro Tools HDX avec son interface HD MADI. Une console Euphonix System 5 SC264 fait les envois vers le système de diffusion via un convertisseur Lynx Aurora 16. Il alimente en analogique, un amplificateur Yamaha XM4180 pour le surround, et en AES, les processeurs de diffusion, Meyer Sound Galiléo 408. La diffusion LCR (Left, Center, Right) est assurée par trois enceintes Meyer Sound Acheron Studio. Les renforts de graves sont composés de deux enceintes Meyer Sound X 800C. Ces cinq enceintes sont placées dans un mur THX [6] situé derrière un écran tissu [7], transonore. Les arrières gauche et droit sont chacun restitués par trois enceintes JBL 3310. Les enregistrements sont réalisés par des microphones DPA 4060 placés dans les oreilles d'une tête Neumann KU80, d'un mannequin Kemar 45BA et dans mes propres oreilles situées à l'emplacement du mixeur. Les micros à électrets sont placés à la limite de l'oreille externe. Ils sont alimentés par l'alimentation fantôme 48V du préamplificateur Euphonix ML530 à l'aide de l'adaptateur DPA DAD 6001. Le signal est converti par un autre convertisseur Lynx Aurora 16, pour être enregistré sur la même session Pro Tools.

6. Un mur THX est un mur plein et massif, traité acoustiquement, dans lequel sont encastrées les enceintes, afin d'éviter les réflexions sur un mur a l'arrière des enceintes, qui, en se mélangeant avec le son direct, génèrent du filtrage en peigne [4].

7. L'écran tissu, assez récent, est plus transparent acoustiquement que l'écran micro-perforé qui est le plus répendu actuellement dans les salles. Il limite également les réflexions entre l'écran et le haut-parleur qui génèrent du filtrage en peigne.

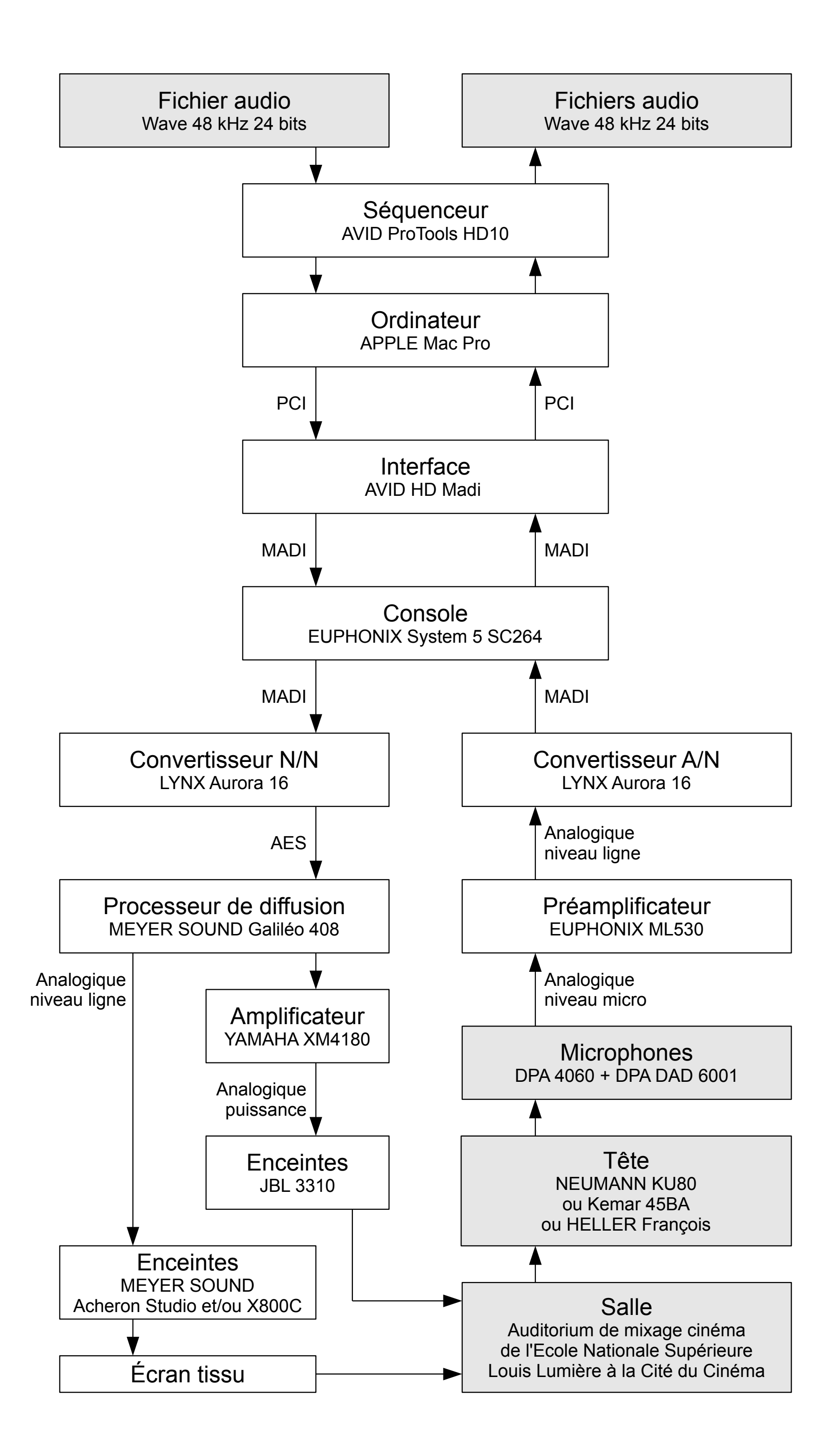


Figure 9 – Chaine de mesure

3.2.3 Réglages effectués

Placement des micros L'expérience de preneurs de son pratiquant l'enregistrement binaural montre que la position du microphone dans l'oreille peut générer plus ou moins de coloration spectrale due à la cavité du pavillon et du conduit auditif [29]. Ici, les microphones miniatures sont placés au raz de l'oreille en sortie de conduit auditif bouché, afin de dégager le capteur des résonances du creux de l'oreille.

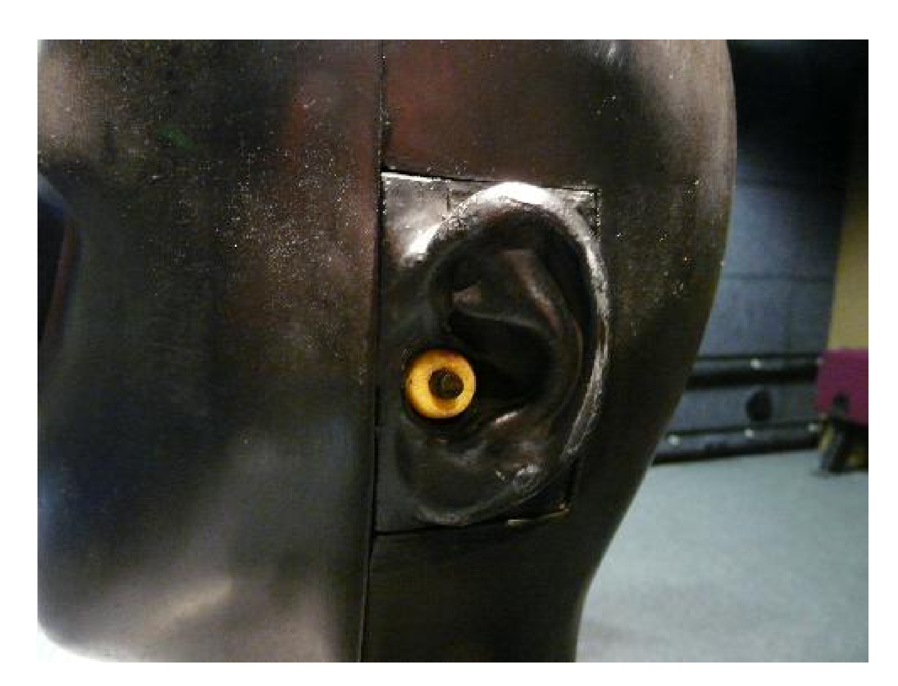


FIGURE 10 – Placement des micros : tête Neumann

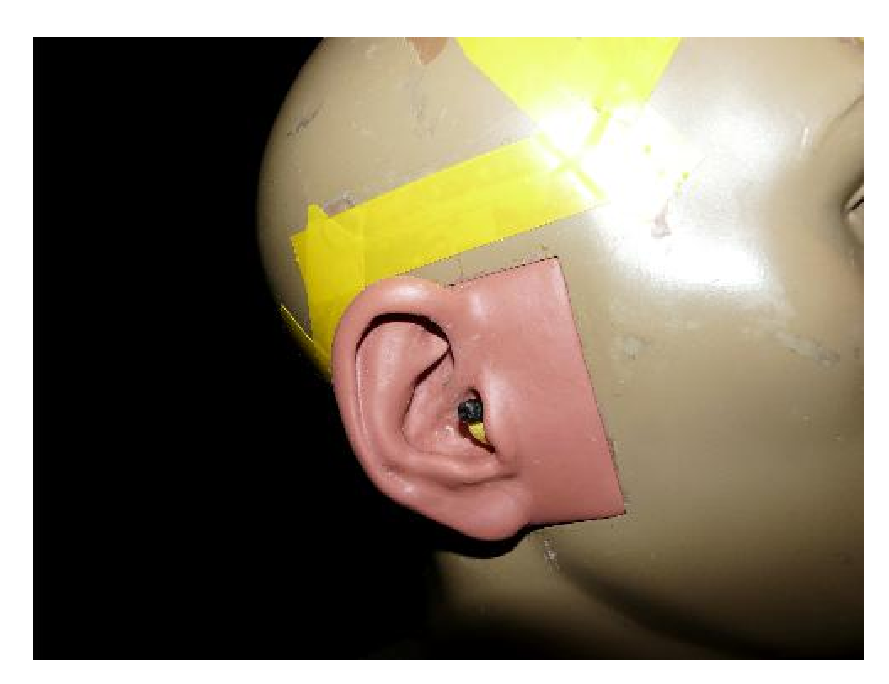


FIGURE 11 – Placement des micros : mannequin Kemar

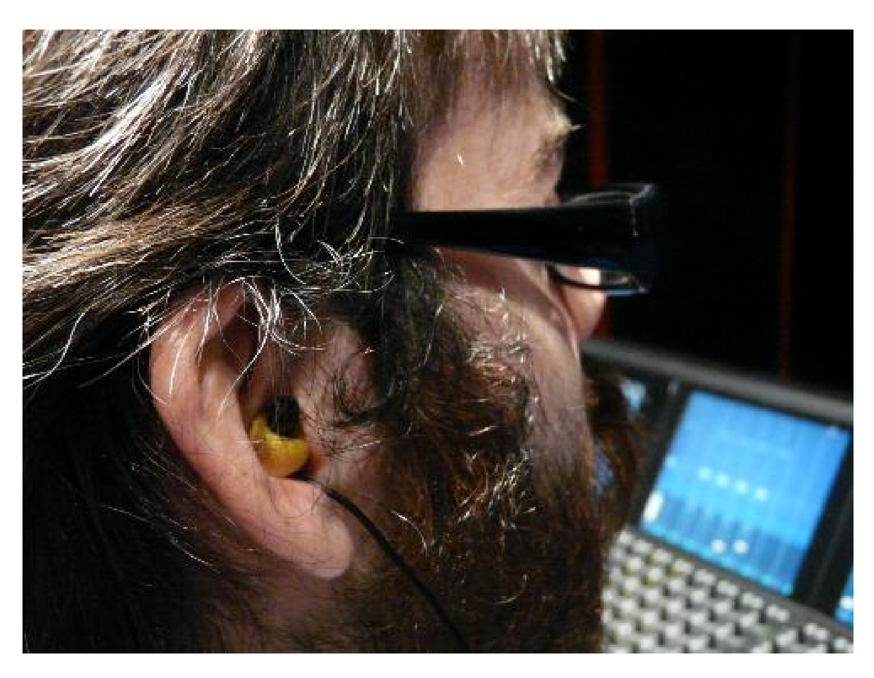


FIGURE 12 – Placement des micros : tête

Traitement du mannequin Les cavités de la tête et du torse du mannequin Kemar sont remplies avec des blocs de mousse pour limiter les résonances internes. Les cavités destinées à accueillir les corps des microphones Neumann de la tête de la même firme sont remplies par du papier pour les mêmes raisons. Les conduits auditifs des oreilles sont systématiquement bouchés par un bouchon de mousse percé dans lequel est passé le microphone miniature.



FIGURE 13 – Tête Neumann KU80 ouverte

Traitement de l'auditorium Les réflecteurs de l'éclairage de l'auditorium, fixés dans le faux plafond suspendu, ont tendance à se mettre à vibrer pour les basses fréquences à fort niveau. Pour diminuer ce phénomène parasite, ces réflecteurs ont été solidarisés plus fermement aux dalles du plafond par des adhésifs, et les stimulus du canal de renfort de graves ont étés baissés de six décibels pour une première passe de mesures, et de douze décibels pour une seconde. Il y a également des dalles et la visserie du faux plafond situées légèrement à l'arrière de la position du mixeur qui entrent en résonance avec certaines basses fréquences. Cette zone est traitée en appliquant une pression sur l'ossature métallique du plafond à l'intersection critique à l'aide d'un pied girafe.

3.2.4 Stimuli

Le stimulus audio utilisé est un fichier wave à la fréquence d'échantillonnage de 48kHz quantifié sur 24 bits. Il est composé de trois sweeps identiques pour s'assurer une sécurité, d'un bruit rose, d'un extrait musical, d'un extrait de parole et d'un extrait d'ambiance. Les différents stimuli sont séparés par six secondes de silence.

Le sweep, balayage fréquentiel, est généré sous Audacity version 2.0.3 [55] , à l'aide du menu « Générer », « Sifflet ». Il a une forme d'onde sinusoïdale, une amplitude d'un quart sur une échelle de zéro à un, une durée de vingt secondes et est constitué d'une interpolation logarithmique entre 20Hz et 20kHz.

L'extrait musical est tiré du premier titre du disque du commerce, Isam (2011) du brésilien Amon Tobin, sous le label Ninja Tune. Il s'agit d'un extrait de trente-trois secondes débutant à une minute, dix secondes et 33333 échantillons. Seule la piste de gauche est conservée pour avoir un stimulus mono. L'acquisition du CD est faite à l'aide de EAC version 1.0 [61] et le passage de 44,1kHz à 48kHz est réalisée par Reaper version 4.32 [52] , qui est réputé pour la qualité de ses conversions de fréquences d'échantillonnage, en sélectionnant « Resample Mode (if needed) : Extreme (512pt HQ Sinc - NON-REALTIME) » dans les options de rendu. Un fade out d'une demi-seconde est généré sous Reaper, et le niveau est baissé de douze décibels sous Audacity.

Le bruit rose est généré sous Audacity version 2.0.3 [55] , à l'aide du menu « Générer », « Bruit ». Il a une amplitude d'un quart, sur une échelle de zéro à un et une durée de dix secondes.

La première moitié de l'ambiance est une ambiance de forêt le matin, tirée de la banque de son Auvidis, disque F « Ambiances de la nature » (1990), piste 01 « Le matin en forêt », en gardant dix secondes de la voie de gauche en partant de neuf secondes et 24000 échantillons. La seconde moitié de l'ambiance est une ambiance de campagne de Provence la nuit, tirée de la piste 10, « La nuit », en gardant dix secondes de la voie de gauche en partant de douze secondes et 45000 échantillons.

L'extrait de parole est issu d'une prise de son en tournage. Il s'agit d'un comédien perché avec un micro hyper-cardioïde Neumann KM185 branché sur un enregistreur Tascam HDP2. Il dure vingt-et-une secondes.

L'ensemble des stimuli module autour de -12dBFS, soit au niveau nominal. Les stimuli envoyés vers le surround sont baissés de six décibels pour éviter les distorsions générées par l'enceinte. De même, les stimuli envoyés vers le LFE sont baissés de six décibels pour une première passe de mesures, et de douze décibels pour une seconde (cf. section 3.2.3, page 34). La démarche est justifiée également par le fait qu'on envoie très rarement des signaux à fort niveau dans les arrières et dans le renfort de graves d'une bande son pour le cinéma. La gestion du niveau se fait directement dans la session ProTools grâce à l'outil « clip gain ».

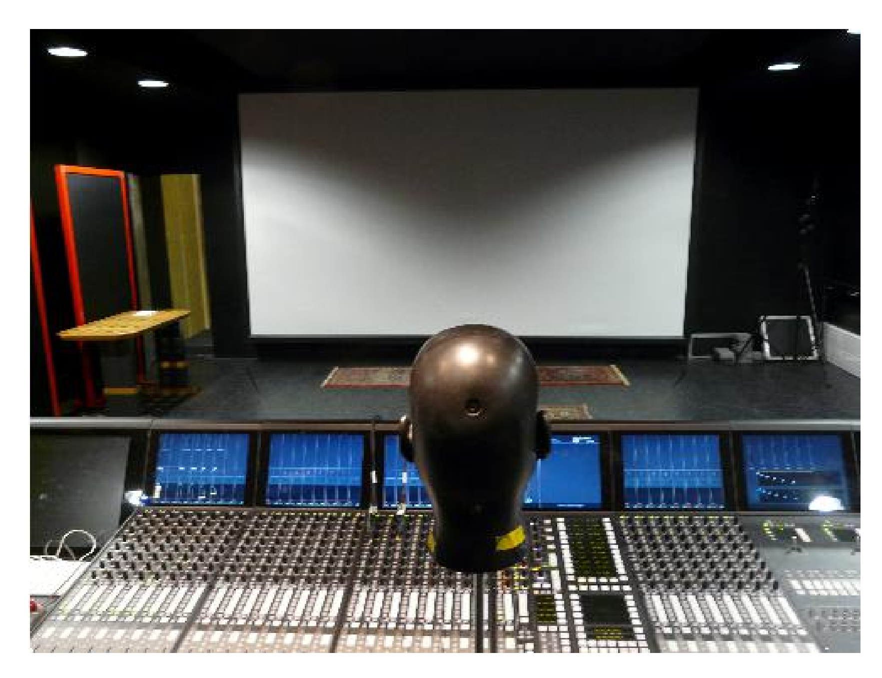


Figure 14 – Configuration Neumann



Figure 15 – Configuration Kemar photo 1

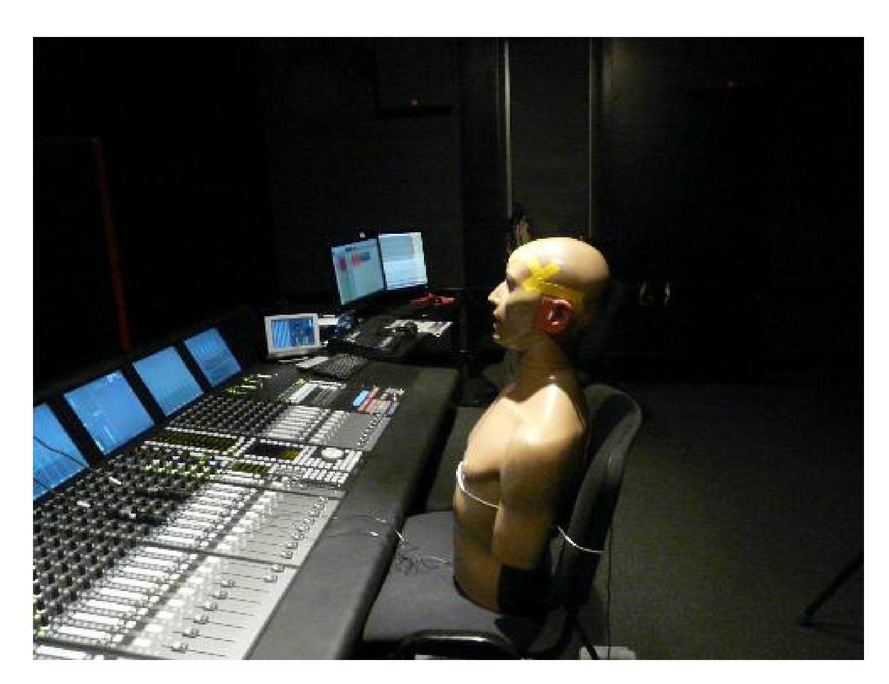


FIGURE 16 – Configuration Kemar photo 2



FIGURE 17 – Configuration Tête

3.3 Obtention des BRIR par déconvolution

L'obtention des réponses impulsionnelles va se faire par déconvolution à partir des sweeps et pour chaque BRIR souhaitées. Ayant enregistré trois sweeps pour chaque configuration on dispose d'une sécurité en cas de parasitage sur l'une des mesures (bruits de couloir ou autres). La déconvolution des fichiers enregistrés par rapport aux fichiers de stimulus est effectuée sous Matlab. On utilise la fonction BruteDecon de Brian Katz du 29 décembre 2003 modifiée le 15 mars 2004 qui est une déconvolution par FFT (cf. 2.4.2, page 25). Les fichiers ainsi générés sont dans un premier temps découpés pour ne garder que 12000 échantillons, soit un quart de seconde, car on est déjà largement dans le bruit de fond et raccourcir les réponses impulsionnelles permet de réduire la charge de calculs lors de la convolution.

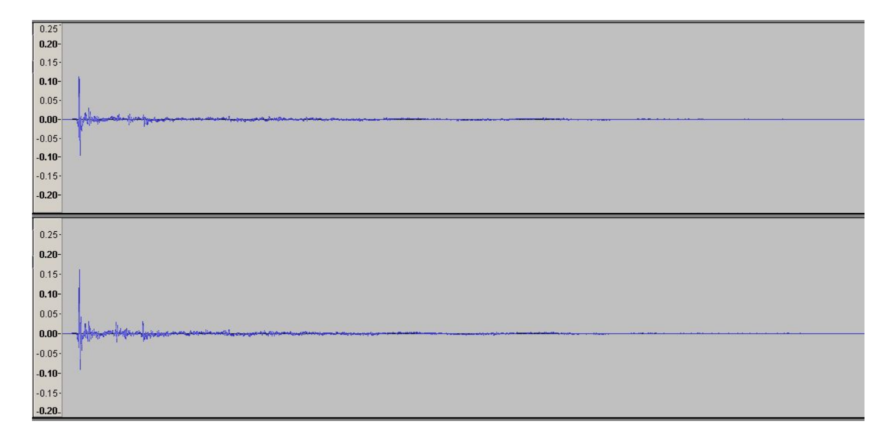


FIGURE 18 – Forme d'onde
pour les réponses impulsionnelles du centre,
oreille gauche en-haut et droite en-bas.

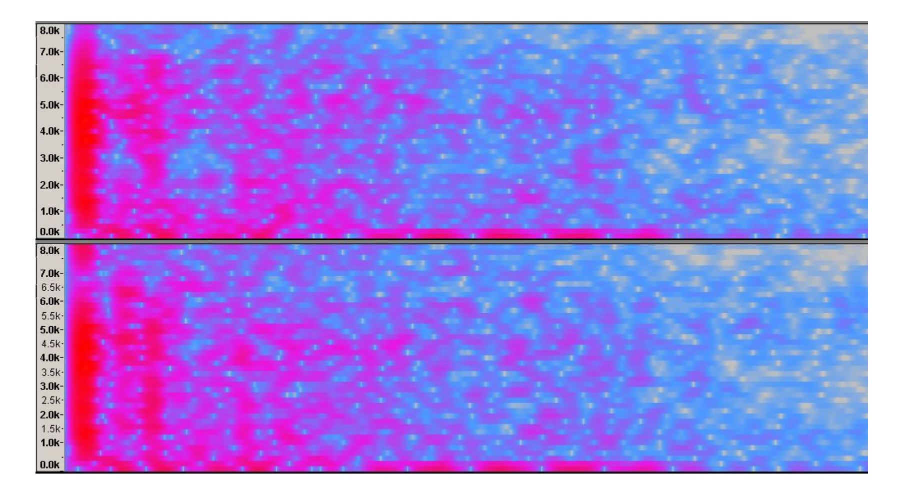


FIGURE 19 – Représentation temps/fréquence ou spectrogramme
pour les réponses impulsionnelles du centre,
oreille gauche en-haut et droite en-bas.

3.4 Validation des BRIR

3.4.1 Méthode

Une fois les réponses impulsionnelles obtenues, on s'attache avant toute chose, à vérifier leur « qualité ». C'est à dire à tester si la convolution d'un son par ces BRIR est semblable à l'écoute d'une prise de son réalisée dans l'auditorium. On effectue le contrôle canal par canal, puis sur un extrait en 5.1. Pour ce faire on aligne dans une session Reaper les stimuli diffusés avec les enregistrements de ces mêmes stimuli. Les sons qui servaient de stimuli ne sont pas dirigés vers le master mais vers un convolueur configuré pour traiter chaque canal par la réponse impulsionnelle correspondante. Cette opération est ici réalisée par la version de démonstration de la réverbération à convolution, Pristine Space de Voxengo, version 1.8 [60]. Elle a été retenue parce qu'elle est gratuite, que l'on peut y entrer ses propres réponses impulsionnelles au format wave et que l'on peut réaliser jusqu'à huit convolutions d'un son mono par un son mono avec une seule occurrence du plug-in. Il suffit ainsi de lancer une de ces réverbérations pour réaliser les six convolutions (L, C, R, Sl, Sr, LFE) pour l'oreille gauche, et une autre pour l'oreille droite. Tous les autres réglages de la réverbération (filtres etc) sont débrayés. Reaper est réglé pour compenser la latence induite par les convolueurs (ici réglée à 32768 échantillons), dans les faits les pistes d'enregistrement qui vont directement au master sont simplement retardées. Enfin, par un jeu de groupes, de bus et de solos il est possible de passer de l'écoute de la version convoluée à la version enregistrée d'un unique clic de souris. Les écoutes sont faites avec un casque Sennheiser HD25 relié directement à la carte son interne d'un ordinateur Dell Vostro 1500. L'annexe E page 63 présente un schéma descriptif.

3.4.2 Observations

Canal par canal Lors d'une comparaison de la diffusion de la prise de son binaurale et de la version binaurale convoluée, on remarque :

- que les différences ne sont pas aisément perceptibles
- que chaque canal est, dans les deux cas, situé dans la partie de l'espace qui lui correspond (l'avant droit est à l'avant droit etc.)
- qu'il y a plus de bruit de fond (dans le haut du spectre notamment) sur la version « prise de son » [8]
- que, dans les deux cas, on retrouve une bande passante assez large (qui descend dans le bas du spectre et monte dans le haut)
- l'impression d'un lieu très légèrement plus diffus dans la version convoluée

Sur un contenu 5.1 Lors d'un passage de la diffusion de la prise de son binaurale à la version binaurale convoluée, on remarque :

8. effectivement, on remarque que plus on allonge la durée du sweep utilisé comme stimulus pour l'aquisition d'une réponse implusionnelle, meilleur sera le rapport signal à bruit lors de la convolution par cette réponse impulsionnelle, jusqu'à surpasser celui de la prise de son [22].

– que les différences sont encore moins perceptibles
– que la spatialisation est assez bien retraduite
– qu'il y a encore plus de bruit de fond sur la version « prise de son »

Des extraits audio correspondants à ces comparaisons seront présentés à la soutenance et sont disponibles sur le support joint au mémoire afin que le lecteur puisse se faire son idée. On peine à distinguer des différences entre les deux versions, au point que l'on pressent que l'on aurait de fortes chances de ne pas pouvoir retrouver l'original avec un test en aveugle. Cela valide les mesures et le procédé de convolution. On peut considérer la mesure comme étant juste et de bonne qualité. On dispose à partir de ce point d'un outil susceptible de créer la simulation d'écoute voulue.

4 Exploitation des BRIR

Il est maintenant nécessaire, pour éprouver le procédé, de l'appliquer à de réelles prises de son multicanal. Il s'agit donc d'enregistrer des ambiances multicanales et d'essayer de voir si l'outil proposé permet de mieux apprécier le rendu spatial final, qu'une réduction stéréo classique. On réalise ces captations avec deux techniques de prises de son différentes, l'une matricée et l'autre non-matricée et dans des lieux divers.

4.1 Captation d'ambiances multicanales

4.1.1 Croix IRT

La première configuration testée est composée des quatre microphones cardioïdes Rode NT5 protégés par des boules Rycotes en configuration de Croix IRT, avec 25cm d'écart entre les capsules montées sur une perchette. L'enregistrement est fait sur un Tascam DR680 contrôlé à l'aide d'un casque Sennheiser HD25. Cette configuration est bien moins pratique que la suivante lors d'une utilisation nomade. En effet il faut gérer quatre câbles, vérifier l'angle des micros et le système est plus lourd. Ces problèmes peuvent êtres contournés avec l'utilisation d'un câble multi-paires et du système Schoeps avec des barres dédiées et des microphones miniatures (la série CCM). Les prises de son on été réalisées dans les lieux suivants :

- Fenêtre sur la ville de Saint-Denis
- Cage d'escalier
- Salon d'une maison avec fenêtre ouverte
- Jardin d'une maison en quartier résidentiel
- Intérieur voiture
- Forêt



FIGURE 20 – Configuration croix IRT photo 1

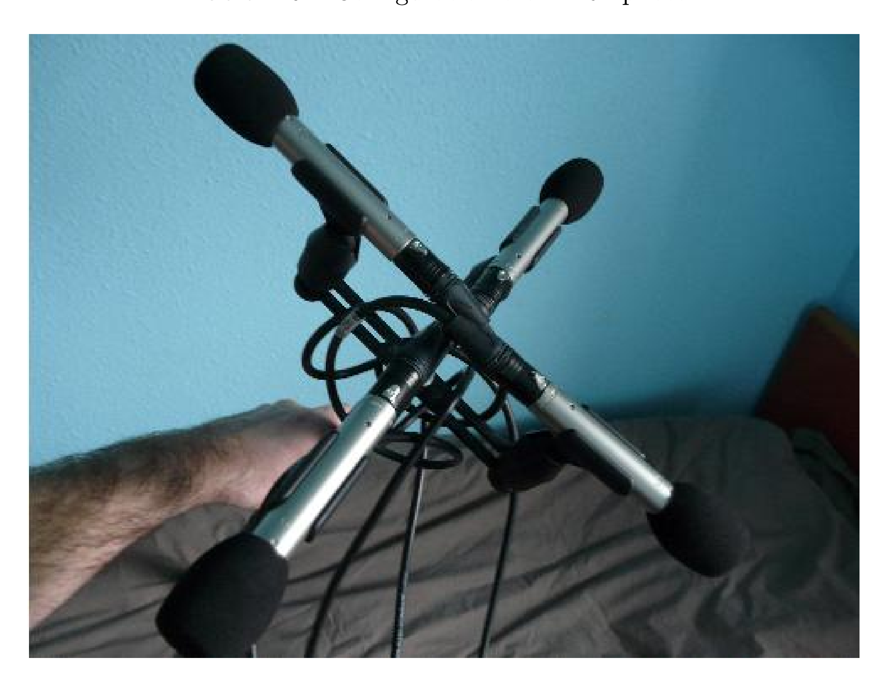


FIGURE 21 – Configuration croix IRT photo 2

4.1.2 Soundfield

La seconde configuration utilisée est destinée à des enregistrements ambisoniques au format B. Elle est composée d'un Soundfield ST450 [47] dans une Rycote, du décodeur Soundfield portable associé, de l'enregistreur Fusion de Zaxcom [48] et d'un casque Sennheiser HD25. C'est une configuration relativement ergonomique et agréable à utiliser au vu des possibilités offertes par les techniques ambisoniques. On a qu'un seul câble à gérer, la suspension est adaptée et fonctionne, la poignée est pratique. Le tout forme un système qui n'est pas trop voyant car compact. Le timbre et la dynamique, canal par canal est tout à fait honorable. Les prises de son on été réalisées dans les lieux suivants :

- Place du 8 mai 1945 à Saint-Denis
- Marché couvert de Saint-Denis
- Basilique de Saint-Denis
- Métro
- Nef de la Cité du Cinéma
- Cafétéria de la Cité du Cinéma
- Hall d'accueil de l'École Louis Lumière
- Gare du Nord
- Escalators à Chatelet
- Place Georges Pompidou

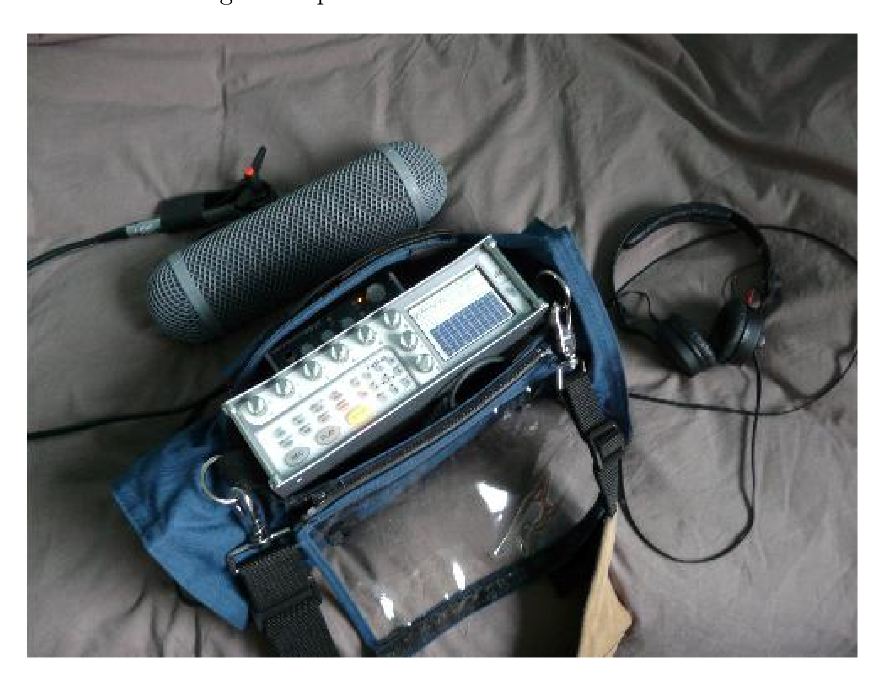


FIGURE 22 – Configuration Soundfield photo 1



FIGURE 23 – Configuration Soundfield photo 2

4.2 Outils logiciels

4.2.1 Décodage du format B

Il y a plusieurs outils qui permettent de décoder le format B dans le format 5.1. Dans le cadre de ce mémoire c'est le plugin-in Harpex B qui a été choisi dans sa version de démonstration pour sa compatibilité avec Windows et VST [9] ainsi que pour sa gratuité. Il semble néanmoins intéressant de présenter rapidement le panorama des outils disponibles selon les recherches effectuées. En effet il permet de se rendre compte de la flexibilité qu'offre l'ambisonie en termes d'ajustements possibles à la post-production.

Surround Zone Soundfield propose son plug-in Surround Zone, qui permet de convertir un signal au format B en un signal 5.1, 6.1, 7.1, 8 canaux ou stéréo. Il offre la possibilité de régler le niveau général d'entrée et de sortie, le niveau de chaque canal de sortie, de définir la position du micro (s'il est suspendu à l'envers ou maintenu à l'horizontale, on permute les composantes pour retomber sur la bonne configuration avec X avant/arrière, Y pour droite/gauche et Z pour haut/bas). Il est également possible de faire des réglages de rotation, d'inclinaison et de zoom

9. VST (sigle de Virtual Studio Technology) est l'un des protocoles pour les plug-in audio existant dans le monde des logiciels. VST a été créé par la société Steinberg pour son logiciel séquenceur Cubase. Les plug-ins VST sont des fichiers compilés ayant pour extension « .vst » sous Mac OS et « .dll » sous Windows.

de l'image multicanale et enfin, on peut déterminer la largeur (en degrés d'angle) de l'avant et de l'arrière, ainsi qu'une directivité de micro virtuel pour la restitution des canaux arrières. Surround Zone existe dans les formats VST, RTAS (pour ProTools), Audio Unit (pour Mac OS) et pour plateforme SADiE.

Harpex B Harpex est une société Norvégienne qui développe des décodeurs du format B vers d'autres formats. Elle commercialise une librairie en C pour l'intégration dans des applications tierces, un plug-in VST, RTAS ou Audio Unit et un lecteur gratuit pour plateforme Linux, Windows ou Mac OS. Le plug-in, dénommé Harpex B est disponible en version de démonstration avec l'intégralité de sa fonctionnalité durant trente jours. Il dispose de cinq modes de fonctionnement : « Binaural » pour une diffusion spatialisée au casque, « Stereo », « Shotgun » pour isoler des sources sonores sur la base de leur localisation dans l'espace ou pour une diffusion sur une distribution 3D de haut-parleurs, « Surround » pour une distribution horizontale de haut-parleurs et « Ambisonics » pour une conversion vers des formats ambisoniques d'ordre supérieur ou des transformations au sein du format B. Dans chaque mode de nombreux « preset » permettent de retrouver rapidement la grande majorité des formats jusqu'à huit canaux. Il propose comme Surround Zone d'ajuster le niveau d'entrée et celui de chaque canal de sortie de réorienter le micro virtuel et de le déplacer pour zoomer. Il permet en plus de retarder les canaux choisis de la durée voulue, de régler une fréquence de filtre passe-bas pour les canaux de renfort de graves. Enfin il propose une visualisation assez originale de l'espace sonore à plat, avec des zones de couleurs, qui correspondent à des bandes de fréquence du spectre audio.

La suite B2X. Daniel Courville a développé toute une série de plug-ins pour la gestion de signaux ambisoniques à l'aide de Sonic Birth, un langage de programmation graphique pour la création de plug-ins VST et Audio Unit pour Mac OS uniquement. Il y a un outil destiné à chaque conversion. Par exemple B2Stereo permet de décoder le format B vers un signal stéréo, B2G permet de décoder le format B vers un 5.1, Tetra2B permet de passer du format A au format B... Il y a également des outils pour traiter des signaux du deuxième, du troisième et du cinquième ordre.

Visual Virtual Mic. VVM est une application « standalone » (autonome) pour Windows, qui accepte du format B et délivre jusqu'à trente-deux signaux correspondant à autant de microphones virtuels. Il est capable, au choix, d'écrire des fichiers sur un disque dur en « batch mode » (pour traitement par lots), ou de lire en temps réel, les trente-deux signaux potentiels, via une carte son.



FIGURE 24 – Harpex B

4.2.2 Convolution

C'est le plug-in Pristine Space de Voxengo qui a été choisi pour la convolution. C'est un outil VST pour Windows qui peut réaliser jusqu'à huit convolutions mono sur huit entrées pour huit sorties. On peut sélectionner un mode sans latence, ou bien fixer la latence voulue. La version de démonstration gratuite dispose de toute la fonctionnalité, avec la sortie coupée pendant un bref instant toutes les vingt-cinq secondes. De nombreux autres convolueurs ont étés trouvés lors des recherches pour ce travail. Ils sont listés et brièvement décrit en annexe (cf. annexe F page 64).

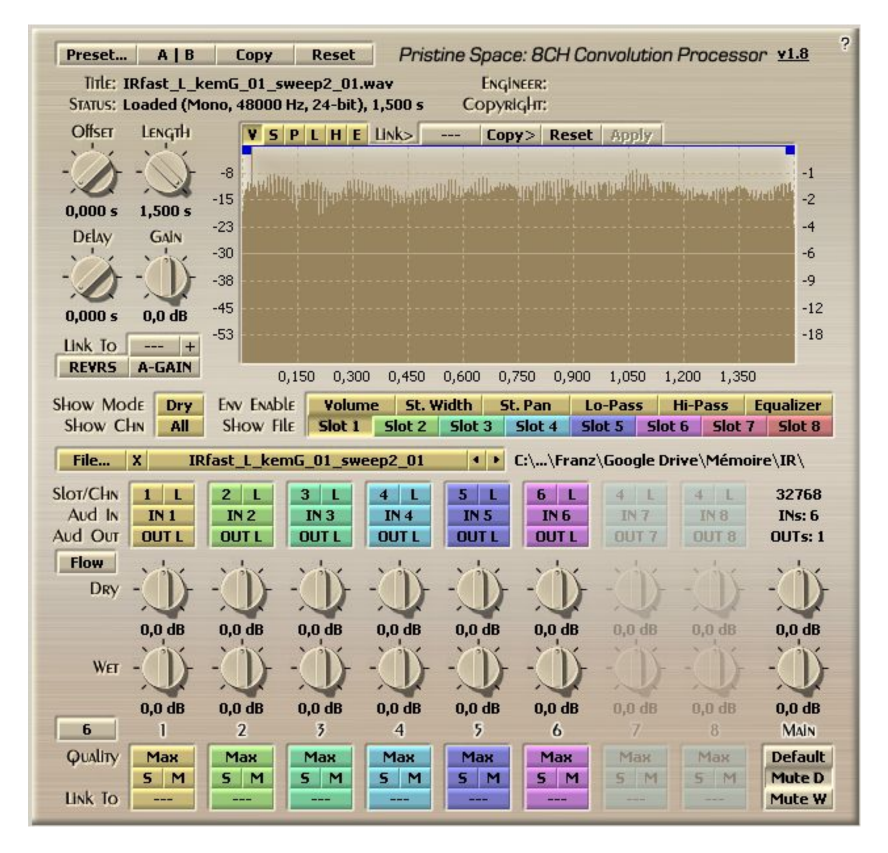


FIGURE 25 – Pristine Space

4.3 Traitement et résultats

Pour tenter de sonder l'apport possible de cet outil de monitoring on souhaite comparer une réduction stéréo, ou « downmix », classique avec la version binaurale proposée, les deux dans le cas d'une écoute au casque comme c'est le cas en situation de tournage. A partir du corpus d'ambiances enregistrées on a donc mis en œuvre deux traitements en parallèle pour disposer d'une version binaurale et d'une version stéréo synchrones, avec la possibilité de passe de l'une à l'autre en temps réel avec une seule commande.

4.3.1 Préparation des extraits audio

Version binaurale On importe les prises de son sur six pistes d'une session Reaper. Les prises de son au Soundfield, décodées avec Harpex B en 5.1, donc sur six pistes, alimentent les six canaux. Les prises de son à la croix IRT sont elles en 4.0. Elles n'alimentent donc que les pistes L, R, Sl, et Sr. Chacune des six pistes de Reaper va alimenter simultanément deux instances du convolueur Pristine. On a donc deux plug-in avec six entrées et six sorties chacun, soit douze convolutions mono simultanées. Les six sorties du premier convolueur sont sommées et envoyées à la voie master gauche tandis que celles du second convolueur sont sommées

et envoyées vers la voie droite. Ainsi, la piste des arrière droite, par exemple, va à la fois alimenter la voie master gauche via une convolution et la voie master droite via une autre convolution. On obtient finalement un contenu binaural, sur deux canaux simulant une diffusion multicanale 5.1.

Version stéréo De manière parallèle, les six pistes L, C, R, Sl, Sr, LFE de Reaper sont également envoyées dans un autre bus master stéréo avec les réglages suivants : le canal de gauche est envoyé dans la voie de gauche, le canal de droite est envoyé dans la voie de droite, le canal central est envoyé dans les deux voies (une atténuation automatique de trois décibels est effectuée de chaque côté pour conserver le niveau), le canal arrière gauche est envoyé à gauche après une atténuation de trois décibels, le canal arrière droit est envoyé à droite après une atténuation de trois décibels, le canal de renfort de graves est envoyé dans les deux voies après une atténuation de trois décibels (ici également une atténuation automatique de trois décibels supplémentaires est effectuée de chaque côté pour conserver ce niveau). Ce traitement correspond à la réduction la plus classique d'un contenu 5.1 vers un flux stéréo.

4.3.2 Écoutes comparatives et bilan

La version binaurale constitue la proposition pour un monitoring pouvant peut-être permettre de mieux estimer le rendu spatial final d'une prise de son multicanal. Lors d'une comparaison entre la diffusion des prises de son via les enceintes virtuelles en 5.1, d'une part, et de la réduction stéréo, d'autre part, on remarque :

- une bien meilleure externalisation sur la version binaurale,
- une variation de timbre sans pouvoir dire quelle version serait la plus « fidèle », celui de la version binaurale est plus resserré,
- une forte correspondance entre la spatialisation entendue par l'écoute binaurale et celle ressentie lors de l'enregistrement (plus par la vue que par l'ouïe compte tenu du caractère fermé du casque utilisé)
- une suppression de l'impression d'espace lors du passage vers la réduction stéréo,
- un souffle qui semble plus présent sur la réduction stéréo.

Ces impressions rendent l'exploitation de cette méthode tout à fait prometteuse… On approche un rendu de spatialisation tel qu'on le recherchait. Pour que le lecteur puisse se faire son avis, des exemples audio sont disponibles sur le support lié au mémoire et seront diffusés lors de la soutenance.

Pour poursuivre, il faudrait aussi procéder à une écoute des ambiances dans l'auditorium afin de vérifier si l'espace à été correctement anticipé. On peut ensuite envisager de définir un protocole de test d'écoute pour valider l'intérêt de cet outil de monitoring, mais ce dernier ne semble pas trivial a mettre en œuvre : on peut imaginer recréer une scène sonore

spatialisée en studio, avec plusieurs enceintes pour les différentes sources qui la composent, pour se placer dans une situation reproductible. Un preneur de son aurait à disposition ses outils classiques de monitoring, ou bien la formule proposée dans ce mémoire, et il s'agirait d'essayer de voir dans quelle mesure ce nouvel outil permettrait d'obtenir ou non un résultat généralement plus satisfaisant à la ré-écoute des prises dans un auditorium de mixage.

4.4 Perspectives

Les solutions présentées ci-après n'ont pu être mises en œuvre dans le cadre de ce mémoire car elles nécessitent de nouveaux développements que le temps imparti ne permettait pas de réaliser. Toutefois il est présenté ici les pistes qui semblent les plus prometteuses et intéressantes.

4.4.1 Portabilité

Le principal intérêt du procédé testé dans ce mémoire réside dans une utilisation nomade sur le terrain. A terme, on pourra envisager que la convolution des différents canaux pour une écoute « binauralisée » se fasse directement dans l'enregistreur qui intègre déjà des unités de traitement numérique du signal. Plusieurs solutions sont envisageables d'ores et déjà, pour réaliser un prototype pour une utilisation portable :

Raspberry Pi Raspberry Pi est une fondation à but caritatif anglaise. Son objectif est de proposer un ordinateur simple et peu cher dans le but de permettre au plus grand nombre d'accéder à l'informatique et d'apprendre à programmer. La fondation a ainsi développé le raspberry Pi (aujourd'hui dans sa version B) qui est un ordinateur à carte unique de la taille d'une carte de crédit fourni sans boîtier, ni alimentation, ni clavier, ni souris, ni écran. Il bénéficie entre autres d'un processeur ARM 700 MHz, de 512 Mo de mémoire vive, d'un port USB 2.0, d'un port HDMI et d'un port Ethernet. Il est suffisamment puissant pour décoder des flux vidéo full HD. Il est alimenté en courant continu 5 Volts. Il permet d'exécuter certaines versions du système d'exploitation libre Linux stocké sur une carte SD [59].

L'environnement Linux n'est malheureusement pas encore compatible avec toutes les interfaces d'entrées/sorties audio. Les interfaces dites « class compliant » sont standardisées et peuvent fonctionner sous Linux sans trop de problèmes. A titre d'exemple, la Fireface UCX de RME possède un mode « class compliant » qui permet jusqu'à huit entrées et nécessite une alimentation en courant continu 9-18 Volts [38].

Pure Data est un langage de programmation graphique open source. Il fonctionne sous Linux, Windows ou Mac OS. Il est capable de réaliser du traitement audio en temps réel. Dans sa version « extended » il intègre une librairie qui contient des objets « FFT » (cf. 2.4.2, page 25) permettant des conversions temps/fréquence avec lesquels il est possible

de faire des convolutions. Avec l'interface audio adaptée le Raspberry Pi devient donc une solution envisageable pour tester ce procédé sur le terrain.

Pandaboard La Pandaboard est un équivalent du Raspberry Pi un peu plus onéreux et encombrant mais intégrant un DSP programmable qui permettrait éventuellement de réaliser les convolutions si le processeur du Raspberry Pi s'avérait d'une puissance insuffisante.

Pure Data sous Androïd Enfin pour une portabilité accrue, il est possible de porter des « patch » Pure Data sur des terminaux Androïd tels qu'un téléphone portable ou une tablette. Mais encore faut-il trouver une interface audio compatibles, comportant au moins six entrées et que le terminal soit à même de traiter en temps réel les entrées, les sorties et les convolutions en temps réel...

On peut également s'intéresser à d'autres formules sur téléphone portable ou tablette graphique intégrant maintenant des puissances de calculs importantes et ayant l'avantage d'être à portée de main de n'importe quel preneur du son.

4.4.2 Prise en compte des mouvements de la tête

Pour rendre plus robuste le procédé de localisation en écoute binaurale, on peut procéder au « Head Tracking » (cf. 2.3.2 page 23) qui consiste à mesurer les mouvements de la tête pour en tenir compte dans la reproduction de la scène donnée à entendre et permettre à l'auditeur d'avoir des mouvements de tête par rapport à l'univers restitué au casque. la formule ambisonique pourrait se révéler être particulièrement intéressante dans ce domaine, parce que les mouvements de rotation de toute la scène sonore est relativement aisé à effectuer (on retrouve par exemple des contrôles dédiés à cette opération dans les décodeurs du format B (cf. 4.2.1 page 44), ce qui permet d'éviter de devoir mettre à jour la convolution par les BRIR.

4.4.3 Pistes alternatives

Malheureusement, une seule solution pour répondre à la question du contrôle en prise de son multicanal mobile a été étudiée dans ce mémoire. Qu'en est-il des casques à multiple transducteurs ? Il existe des modèles, conçus pour le jeu vidéo qui regroupent des transducteurs autour de chacune des deux oreilles (deux exemples sont donnés en [63] et [64]), et d'autres qui appliquent des transducteurs directement sur le pourtour du crâne. La possibilité qu'offre Harpex notamment, d'un décodage direct du format B vers binaural n'a pas été abordée. Dans son mémoire, Florent Castellani [14] a également utilisé le programme AudioStage de Longcat pour un passage de l'ambisonique au binaural. Enfin on peut également envisager un contrôle via une représentation graphique, à l'image de la visualisation par fréquences de Harpex B (cf. 4.2.1 page 45). Après passage par des algorithmes de séparation

de sources, on peut imaginer faire apparaitre ces sources distinctes les unes des autres, dans une représentation en deux ou trois dimensions de la scène sonore captée...

Conclusion

Dans ce mémoire, on s'est intéressé à la problématique du contrôle d'écoute de prise de son multicanal dans un contexte de tournage. Le principe des enceintes virtuelles à été mis en œuvre, en prenant soin de n'utiliser que des outils facile d'accès pour un preneur de son au cinéma : avec la possibilité de passer par des auditoriums de mixage, les micros DPA qui sont utilisés couramment en tournage comme micros cravates, les versions des logiciels utilisées sont gratuits (Reaper, Audacity, Pristine Space Demo, Harpex B Demo. . .). Les résultats sont encourageants et apportent des éléments jusque-là non éprouvés en situation de tournage. La voie de l'allègement de la solution de simulation d'enceintes virtuelles, pour la rendre portable est sans aucun doute une piste à creuser. De plus, avec les ressources en calculs disponibles sur les enregistreurs allant en s'accroissant et la similitude de plus en plus marquée avec l'informatique (l'enregistreur 4minX, par exemple, exploite une distribution Linux [34]) des moteurs de convolution seront, probablement prochainement, en mesure d'être implémentés dans les enregistreurs mobiles de terrain. Enfin, l'utilisation du « head tracking » est une option à explorer, car elle est à même d'augmenter la robustesse du procédé quant aux inversions avant/arrière. . . mais de là à la voir se répandre pour une utilisation quotidienne en tournage il reste des murs à franchir.

Bibliographie

Livres

[1] Jens Blauert. *Spatial Hearing. The psychophysics of human sound localization.* Trad. par John S. Allen. Revised. The MIT Press, 1996. isbn : 0-262-02413-6.

[2] Francis Rumsey. *Spatial Audio.* Music Technology. Focal Press, 2001. isbn : 0-240-51623-0.

[3] Vincent Magnier. *Guide de la prise de son pour l'image. Reportage, documentaire, fiction en radio et télévision.* 2e éd. Dunod, 2011. isbn : 978-2-10-053814-0.

[4] Collectif. *Dictionnaire encyclopédique du son.* Sous la dir. de Pierre-Louis de Nanteuil. Dunod, 2008. isbn : 978-2-10-005979-9.

[5] Laurent Millot. *Traitement du signal audiovisuel. Applications avec Pure Data.* Audio-Photo-Vidéo. Dunod, École Nationale Supérieure Louis Lumière, 2008. isbn : 978-2-10-050688-0.

[6] Collectif. *Le Livre des techniques du son. Tome 1 - Notions fondamentales.* Sous la dir. de Denis Mercier. 4e éd. Audio-Photo-Vidéo. Dunod, 2010. isbn : 978-2-10-054919-1.

[7] Collectif. *Le Livre des techniques du son. Tome 3 - L'exploitation.* Sous la dir. de Denis Mercier. 3e éd. Audio-Photo-Vidéo. Dunod, 2007. isbn : 978-2-10-049651-8.

[8] Alain Besse. *Salles de projection, salles de cinéma. Conception, réalisation, exploitation.* Audio-Photo-Vidéo. Dunod, 2007. isbn : 978-2-10-049673-0.

[9] Michael Williams. *Microphone arrays for stereo and multichannel sound recording. Volume 1 - The training of the ear, the operational approach to microphone array sound recording, tutorial on microphones.* Editrice Il Rostro, 2004. isbn : 88-7365-073-2.

[10] Maurice Bellanger. *Traitement numérique du signal. Théorie et pratique.* 7e éd. Dunod, 2004. isbn : 2-10-048634-9.

[11] Raymond Murray Schafer. *Le Paysage sonore. Le monde comme musique.* Trad. par Sylvette Gleize. Domaine Sauvage. Wildproject, 2010. isbn : 978-2-918490-05-0.

Thèses et mémoires

[12] Alan Blum. « Étude de la plasticité du système auditif en localisation sonore. Application au problème de l'individualisation en synthèse binaurale ». rapport de DEA. Université d'Aix-Marseille 2, 2003.

[13] Pierre BOMPY. « Contrôle au casque d'une prise de son multicanale ». Sous la dir. de Benjamin BERNARD et Mohammed ELLIQ. mémoire de fin d'études. École Nationale Supérieure Louis Lumière, 2008. URL : http://www.ens-louis-lumiere.fr/fileadmin/recherche/Bompy-son-2008-mem.pdf.

[14] Florent CASTELLANI. « Adaptation du mixage cinéma à l'écoute au casque ». Sous la dir. de Benjamin BERNARD et Claude GAZEAU. mémoire de fin d'études. École Nationale Supérieure Louis Lumière, 2011.

[15] Aurore CLEMENT. « La prise de son multicanale en cinéma animalier ». Sous la dir. de Philippe BARBEAU et Christian CANONVILLE. mémoire de fin d'études. École Nationale Supérieure Louis Lumière, 2007.

[16] Jérôme DANIEL. « Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia ». thèse de doctorat. Université Pierre et Marie Curie - Paris 6, 2001. URL : http://gyronymo.free.fr/audio3D/downloads/These-original-version.zip.

[17] Antoine GOUNY. « Les HRTF appliquées au retour de scène par « in ear monitor » ». Sous la dir. de Olivier WARUSFEL et Thierry CODUYS. mémoire de fin d'études. École Nationale Supérieure Louis Lumière, 2010. URL : http://www.ens-louis-lumiere.fr/fileadmin/recherche/Memoire_Gouny_2010.pdf.

[18] Pierre GUILLON. « Individualisation des indices spectraux pour la synthèse binaurale : recherche et exploitation des similarités interindividuelles pour l'adaptation ou la reconstruction de HRTF ». thèse de doctorat. Université du Maine, 2009.

[19] Véronique LARCHER. « Techniques de spatialisation des sons pour la réalité virtuelle ». thèse de doctorat. Université Pierre et Marie Curie - Paris 6, 2001. URL : http://articles.ircam.fr/textes/Larcher01a/index.pdf.

[20] Arnaud ROLLAND. « Comparaison entre l'écoute au casque et l'écoute sur haut-parleurs en stéréophonie ». Sous la dir. de Véronique LARCHER. mémoire de fin d'études. École Nationale Supérieure Louis Lumière, 1998.

Articles scientifiques

[21] Federic L. WIGHTMAN et Doris J. KISTLER. « Resolution of front-back ambiguity in spacial hearing by listener and source movement ». Dans : *Acoustical Society of America* (1999). paper 2841. URL : http://asadl.org/jasa/resource/1/jasman/v105/i5/p2841_s1?isAuthorized=no.

[22] Angelo FARINA. « Simultaneous measurement of impulse response and distortion with a swept-sine technique ». Dans : *Audio Engineering Society* (2000). Convention Paper 5093. URL : http://www.aes.org/e-lib/browse.cfm?elib=10211.

[23] Kangeun LEE, Changyong SON et Dohyung KIM. « Low complexity binaural rendering for multi-channel sound ». Dans : *Audio Engineering Society* (2009). Convention Paper 7687. URL : http://www.aes.org/e-lib/browse.cfm?elib=14883.

[24] Paolo MARTIGNON, Andrea AZZALI, Densil CABRERA, Andrea CAPRA et Angelo FARINA. « Reproduction of auditorium spatial impression with binaural and stereophonic sound systems ». Dans : *Audio Engineering Society* (2005). Convention Paper 6485. URL : http://www.aes.org/e-lib/browse.cfm?elib=13201.

[25] Renato PELLEGRINI, Clemens KUHN et Mario GEBHARDT. « Headphones technology for surround sound monitoring - A virtual 5.1 listening room ». Dans : *Audio Engineering Society* (2007). Convention Paper 7068. URL : http://www.aes.org/e-lib/browse.cfm?elib=14053.

Cours et livres blancs

[26] Etienne HENDRICKS. *Méthodologie des tests.* cours de 3e année en section Son. École Nationale Supérieure Louis Lumière, 2012.

[27] Jean-Marc L'HÔTEL. *La captation de son multicanal.* « Les cahiers techniques de Tapages ». livre blanc. Tapages, 2008. URL : http://www.afsi.eu/download/sites/default/files/uploads/MulticanalLivreBlanc2008.pdf.

[28] OMNIHEAD. *Le livre blanc de la prise de son binaurale.* livre blanc. Version 1.2. 2003. URL : http://www.univ-brest.fr/mstis/omnihead/accueil.htm.

[29] Pascal RUEFF, Antoine SASAL et Julien Le VU. *Session d'expériences chez France Télécom R&D.* rapport. OmniHead, 2002.

[30] SCHOEPS. *Surround recording techniques.* livre blanc. 2006. URL : http://www.schoeps.de/documents/SCHOEPS_surround-brochure.pdf.

[31] Pierre-Antoine SIGNORET. *Le Son multicanal.* cours de la Formation Supérieure aux Métiers du Son. Version 2.0. Conservatoire National Supérieur de Musique et de Danse de Paris, 2012. URL : http://l.g.s.free.fr/medias/cnsm/Le%20son%20multicanal%20FSMS%20V2_S.pdf.

[32] TRINNOV AUDIO. *La prise de son 5.0 en haute résolution spatiale.* livre blanc. URL : http://www.trinnov.com/wp-content/uploads/downloads/Trinnov-hsr-5.0-fr.pdf.

Documentation technique

[33] AATON. *Aaton Cantar-X1&2.* Version 2.44. 2013. URL : http://www.aaton.com/files/cantar_manual-244r15.pdf.

[34] AETA. *4MinX manuel utilisateur.* 2013. URL : http://www.aeta-audio.com/fileadmin/downloads/pdf_fr/manuel_4MinX.pdf.

[35] DPA. *DPA 5100 user's manual.* URL : http://www.dpamicrophones.com/sitecore/shell/Controls/Rich%20Text%20Editor/~/media/PDF/Download/Users%20Manuals/5100-manual.pdf.

[36] Nagra. *Manuel d'utilisation Nagra-VI.* Version 3.00. 2010. URL : http://www.nagraaudio.com/files/Manuel_NAGRA_VI_Fr_V3.pdf.

[37] Nagra. *Bulletin de software Nagra-VI.* Version 3.10. 2012. URL : http://www.nagraaudio.com/files/NagraVISoftwarebulletinV3_10FR.pdf.

[38] RME. *User's guide fireface UCX.* URL : http://www.rme-audio.de/download/fface_ucx_e.pdf.

[39] Schoeps. *La shpère microphonique KFM6. Conseils d'utilisation.* URL : http://www.schoeps.de/documents/Schoeps_Manual_KFM6-F.pdf.

[40] Sonosax. *Sonosax SX-ST / SX-VT mode d'emploi.* 2010. URL : http://sonosax.ch/mixers/sxst/201006-SX-ST_VT-UMF.pdf.

[41] Sonosax. *Sonosax SX62R mode d'emploi.* 2011. URL : http://sonosax.ch/recorders/sx62r/sx62R-umfr.pdf.

[42] Sonosax. *Sonosax SX-R4 User Interface.* Version 4.2. 2012. URL : http://sonosax.ch/recorders/sxr4/sxr4_user%20interface_v4.2_eng.pdf.

[43] Sound Devices. *664 Portable production mixer and recorder user guide and technical information.* Version 1.04. URL : http://www.sounddevices.com/download/guides/664_en.pdf.

[44] Sound Devices. *744T High resolution digital audio recorder with time code user guide and technical information.* Version 2.67. URL : http://www.sounddevices.com/download/guides/744t_en.pdf.

[45] Sound Devices. *788T High resolution digital audio recorder with time code user guide and technical information for 788T and 788T-SSD recorders.* Version 2.19. URL : http://www.sounddevices.com/download/guides/788t_en.pdf.

[46] Soundfield. *SPS200 software controlled microphone user guide.* Version 2.0. URL : http://www.soundfield.com/downloads/sps200-manual.pdf.

[47] Soundfield. *ST450 portable microphone system user guide.* Version 2.0. URL : http://www.soundfield.com/downloads/st450_manual.pdf.

[48] Zaxcom. *Fusion 10/12 user's manual.* Version 7.08. 2010. URL : http://www.zaxcom.com/manuals/Zaxcom%20Fusion%20User%20Manual%20-%20March%202010%20Revision.pdf.

[49] Zaxcom. *Deva 16 & Deva 5.8 user's manual.* Version 7.52. 2012. URL : http://www.zaxcom.com/manuals/Zaxcom%20Deva%20User%20Manual%20Release%20-%20June%202012%20Revision.pdf.

[50] Zaxcom. *Nomad user's guide*. Version 4.45. 2013. url : http://www.zaxcom.com/manuals/Zaxcom%20Nomad%20User%20Manual%20-%20January%202013%20Revision.pdf.

[65] Trinnov Audio. *SRP Surround Recording Platform. Le premier microphone multicanal à Haute Résolution Spatiale*. brochure. url : http://www.trinnov.com/wp-content/uploads/downloads/SRP_FR_A4_Jan2007.pdf.

[66] Helmut Wittek. *Schoeps surround microphones*. brochure. Schoeps, 2012. url : http://www.schoeps.de/documents/SCHOEPS_Surround_Wittek_2012_20120417185740.pdf.

Ressources internet

[51] AFSI. *Forum*. url : http://www.afsi.eu/forum.

[52] Cockos. *Reaper*. Version 4.32. 2013. url : http://www.reaper.fm.

[53] Collectif. *Mac The Ripper*. Version 2.6.6. 2002. url : http://www.macupdate.com/download.php/14414/mactheripper266.dmg.zip/.

[54] SourceForge Collectif. *mAC3dec*. Version 1.6.1. 2004. url : http://mac3dec.sourceforge.net.

[55] SourceForge Collectif. *Audacity*. Version 2.0.3. 2013. url : http://audacity.sourceforge.net.

[56] Holophone. *Surround mics*. url : http://www.holophone.com/Holophone_Surround_Sound_Microphones_s/1847.htm.

[57] mh acoustics. *The eigenmike® microphone array*. url : http://www.mhacoustics.com/mh_acoustics/Eigenmike_microphone_array.html.

[58] Perchman.com. *Forum*. url : http://www.perchman.com/forum.php.

[59] Raspberry Pi. url : http://www.raspberrypi.org.

[60] Aleksey Vaneev. *Pristine Space*. Version 1.8. Voxengo. url : http://www.voxengo.com/product/pspace/.

[61] Andre Wiethoff. *Exact Audio Copy*. Version 1.0. 2011. url : http://www.exactaudiocopy.de.

[62] Mike Williams. *SOS MMAD*. Version 0.4. 2009. url : http://www.mmad.info.

[63] Razer. *Tiamat*. url : http://www.razerzone.com/gaming-audio/razer-tiamat-71.

[64] Tritton Audio. *Pro +*. url : http://www.trittonaudio.com/proplus.

Table des illustrations

1 Croix IRT 10
2 Double ORTF 10
3 Carré Hamasaki 11
4 Holophone H2 12
5 OCT surround 12
6 Schoeps KFM 360 14
7 Différences interaurales de temps 20
8 Principe des enceintes virtuelles 29
9 Chaine de mesure 31
10 Placement des micros : tête Neumann 32
11 Placement des micros : mannequin Kemar 33
12 Placement des micros : tête 33
13 Tête Neumann KU80 ouverte 34
14 Configuration Neumann 36
15 Configuration Kemar photo 1 36
16 Configuration Kemar photo 2 37
17 Configuration Tête 37
18 Forme d'onde 38
19 Représentation temps/fréquence ou spectrogramme 38
20 Configuration croix IRT photo 1 42
21 Configuration croix IRT photo 2 42
22 Configuration Soundfield photo 1 43
23 Configuration Soundfield photo 2 44
24 Harpex B 46
25 Pristine Space 47
26 Plan de l'auditorium 60
27 Courbe ISO X 61
28 Comparaison entre convolution et prise de son 63

Annexes

A Détails sur l'expérience OH2

Le réglage du système de diffusion 5.1 est réalisé selon les instructions du document « Multichannel Monitoring Tutorial Booklet » édité par Yamaha. Le système choisi est le suivant : deux Génélec 1037 pour la stéréo avant, une 1031 pour le centre, deux 1031 pour la stéréo arrière. Le grave est pris en charge par deux caissons encastrés, réalisés par France Télécom R & D, équipés de 38cm PHL octogonaux et poussés par un amplificateur Klotz PA900. Les enceintes sont posées sur un cercle de rayon = 220 cm, sur pied ; les hauts parleurs d'aigus sont à une hauteur de 118 cm du sol. Les enceintes L et R forment un angle de 60 ° et les enceintes LS et RS forment chacune un angle de 125 ° avec l'enceinte du centre.

Chaque enceinte est mesurée avec une rampe 20Hz-20kHz depuis le point d'écoute à l'aide d'AudioToolBox, en 1/3 d'octave. Les résultats sont compensés par les égaliseurs 1/3 d'octave de la console Yamaha DM2000. Une pondération à « l'oreille » finalise l'égalisation, notamment pour le canal central dont toute la moitié supérieure du spectre est amplifiée par l'égalisation et paraît très excessive à l'écoute. La pression sonore est calibrée. Chaque enceinte fournit 85dBSPL en mode slow, pondéré A.

La chaîne suivante est mise en œuvre :
- Lecteur DVD CGV AX201, équipé de 6 sorties indépendantes niveau ligne asymétrique
- console Yamaha DM2000
  (avec 6 AD 24/96 pour le DVD, 2 AD 24/96 pour l'OH2, 6 DA 24/96 pour la diffusion et interface ADAT)
- ProTools, interface ADAT
- microphones DPA 4060

Un DVDR contenant quatre extraits de films commerciaux a été préparé. Les extraits choisis sont « Matrix », scène de bataille dans le hall, « Danse avec les loups », scène d'ouverture.

Différentes prises de son sont faites avec un mannequin Kemar 45BA rempli avec un volume plus ou moins important de particules de caoutchouc, pour bloquer les résonances internes. Des prises de son sont également réalisées avec la tête de certains des expérimentateurs.

B Plan de l'auditorium de l'ENSLL

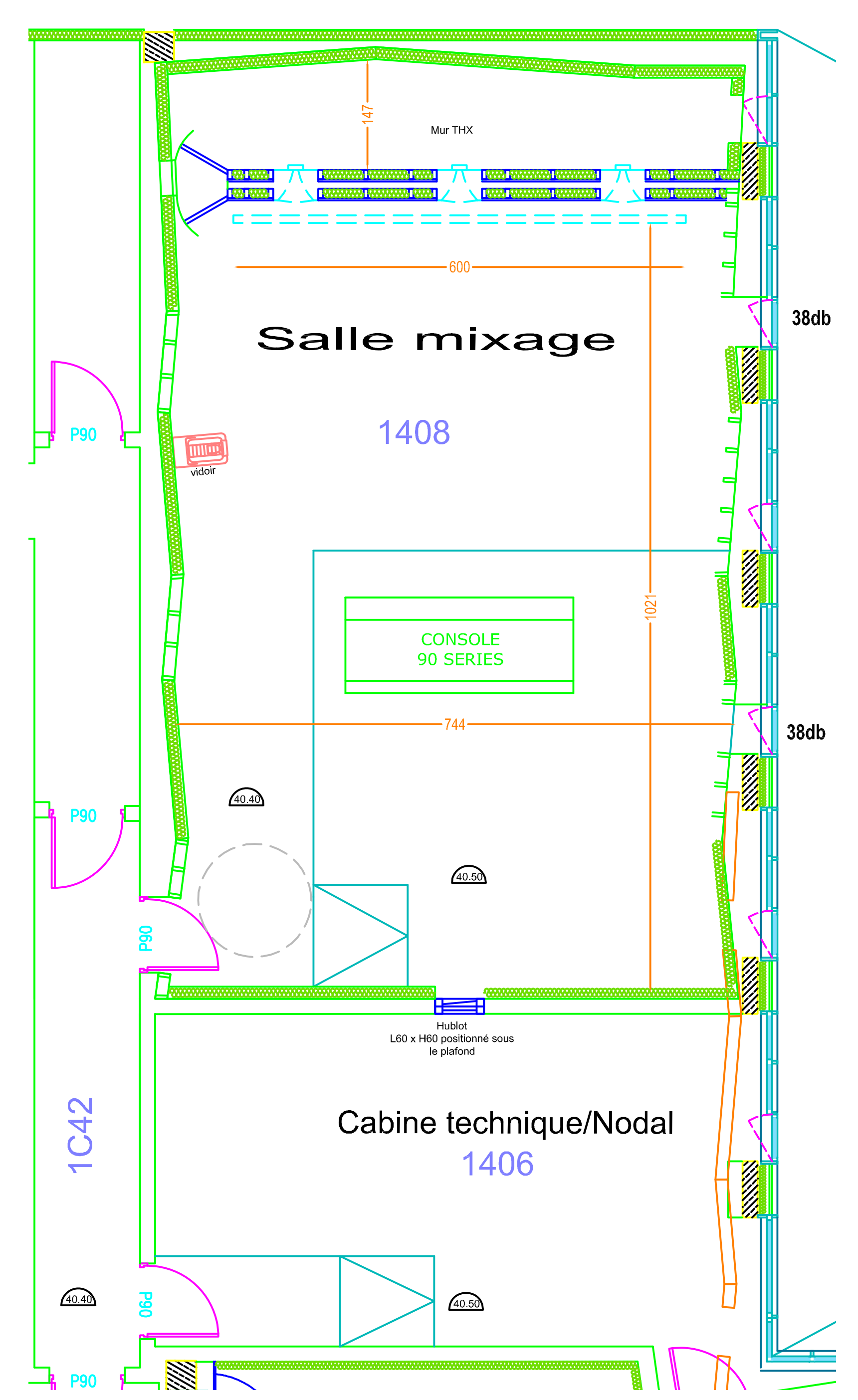


FIGURE 26 – Plan de l'auditorium
Hauteur sous plafond standard de 2,70m

C Norme de diffusion sonore au cinéma

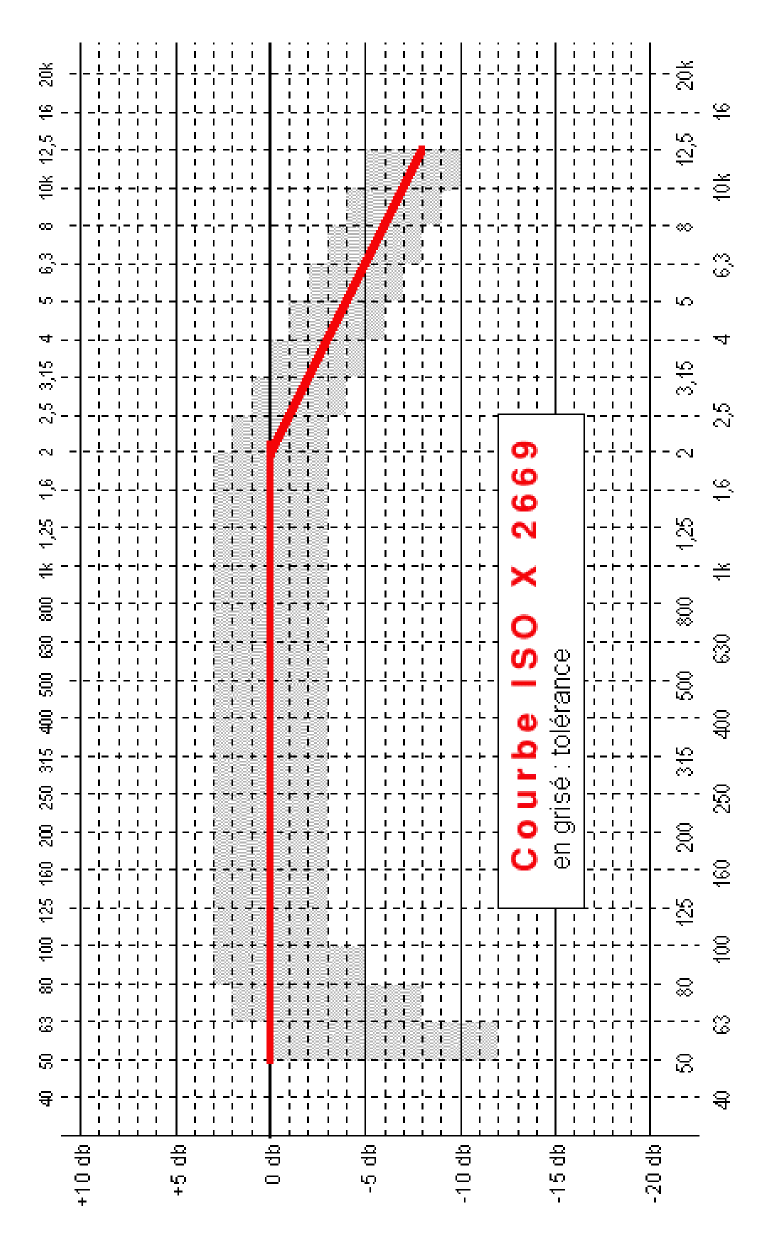


Figure 27 – Courbe ISO X

D Script Matlab pour la déconvolution

```
function ir = BruteDecon(stim,resp,OPT)
% ir = BruteDecon(stim,resp,OPT) Deconvolve two signals using simple FFT
%
% Deconvolve STIMulus from RESPonse of a system
% Created due to lack of success using DECONV
%
% OPT 'fast' uses power of 2 FFT legnth (zero padding) rather than true
% length (seems to cause some minor variations in IR (need to verify if
% significant)
%
% Brian FG Katz
% 29-dec-03
%
% 15-mars-04
% Modified to allow for multichannel RESP using same STIM

if nargin < 3, OPT = 'slow' ; end

if OPT == 'fast'
	% find power of 2 just longer than signal (for the speed of it)
	pfft = ceil(log(length(resp))/log(2)) ;
	nfft = 2^pfft ;
elseif OPT == 'slow'
	nfft = length(resp) ;
else
	error('bad OPT string') ;
end

STIM = fft(stim,nfft) ;
RESP = fft(resp,nfft) ;

for loop = 1 :size(resp,2),
	IR( :,loop) = RESP( :,loop)./STIM ;
	% dont know the syntax to do this without a loop
	IR(1,loop) = 0 ;
	% Set DC component to 0
end % loop

ir = real(ifft(IR,nfft,1)) ;
% ir = real(ir(1 :length(stim), :)) ; % remove excess length from fft padding
% IR returned is cropped to lenght of stimulus
% Need to verify that this is not a bad idea

return
```

E Validation des BRIR : routage du signal

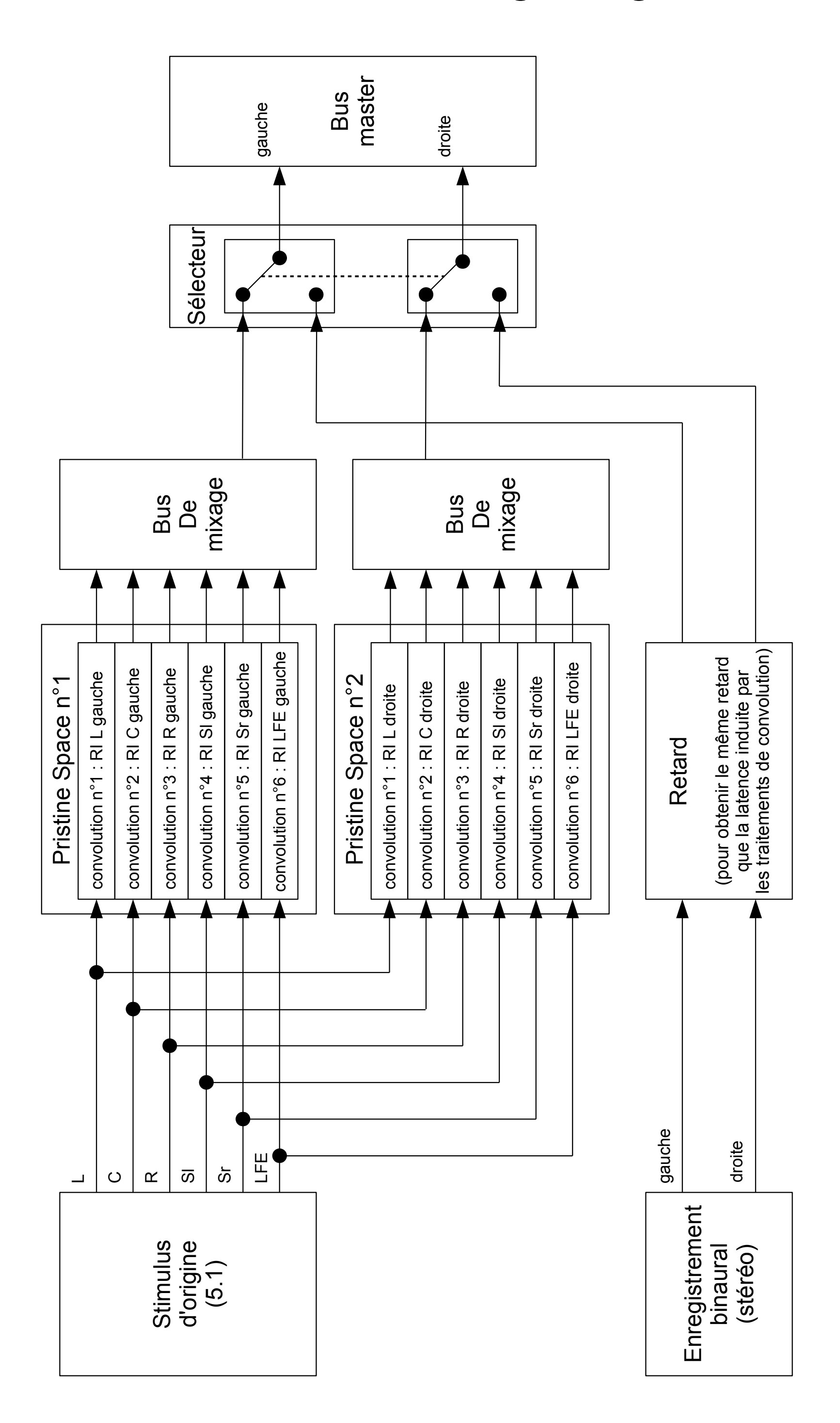


Figure 28 – Comparaison entre convolution et prise de son

F Outils logiciels de convolution

Il est présenté ci-après une liste d'outils issue des recherches préliminaires pouvant permettre une opération de convolution. L'outil qui aura servi pour la partie pratique est le plug-in Pristine Space.

F.1 Plug-ins gratuits

SIR1 : un VST « freeware » de convolution a latence fixe de 8960 échantillons pour Windows

Sound Revolverb Lite : un « abandonware [10] » de l'éditeur Redstate, VST pour Windows avec un mode de fonctionnement sans latence.

Convoboy Boy Lite : de Elevayta, un VST « freeware » de convolution a latence fixe de 4096 échantillons pour Windows

Reverberate LE : est un VST « donationware [11] » pour Windows de LiquidSonics qui fonctionne avec une latence de 0 à 8192 échantillons et qui présente la particularité de pouvoir calculer avec les ressources de la carte vidéo pour alléger la charge du processeur. (fonctionne avec les cartes NVIDIA GeForce 8 series et versions ultérieures avec une latence de 512 à 8192 échantillons)

keFIR : est un VST « freeware » pour Windows conçu à la base pour réaliser des traitements de filtrage à réponse impulsionnelle finie (FIR) sans latence. Il est d'une grande simplicité : quatre boutons, pour choisir la réponse impulsionnelle, régler sa longueur, le gain de sortie et le rapport direct/retour.

ReaVerb : de Cockos, est la réverbération à convolution que l'on trouve par défaut dans le séquenceur Reaper du même éditeur.

Boogex : de Voxengo est un « freeware » VST pour Windows qui sert avant tout à modéliser un ampli guitare, avec un contrôle de gain, un égaliseur, des contrôles de tonalité, de phase. . .et un étage de simulation de baffle utilisant la convolution avec une latence de 96 échantillons.

10. Un abandonware est un logiciel considéré comme abandonné sous prétexte qu'il n'est plus supporté ni exploité commercialement par son éditeur ou ayant droit, si bien que certains utilisateurs prennent des libertés par rapport à la législation sur la propriété intellectuelle pour l'utiliser, le reproduire et le partager gratuitement bien qu'il soit encore protégé. La notion d'abandonware n'a pas de base légale mais peut se défendre devant un tribunal comme un droit coutumier du fait de l'obsolescence rapide d'un logiciel.

11. Un donationware est un freeware dont la rétribution est laissée à l'appréciation de l'utilisateur.

Pristine Space : de Voxengo également est un outil VST pour Windows qui peut réaliser jusqu'à huit convolutions mono sur huit entrées pour huit sorties avec un mode sans latence possible. La version de démonstration gratuite dispose de toute la fonctionnalité, avec la sortie coupée pendant un bref instant toutes les vingt-cinq secondes.

F.2 Plug-ins payants

SIR2 : un plug-in VST ou AudioUnit pour Windows avec une latence de 0 à 1024 échantillons et un fonctionnement « true stereo [12] » à partir de quatre réponses impulsionnelles mono.

IR1A : de Mellowmuse, un plug-in VST, RTAS ou Audio Unit pour Windows ou Mac OS, qui fonctionne sans latence avec des réponses impulsionnelles jusqu'à dix secondes.

Reverence de Steinberg, est la réverbération à convolution que l'on trouve par défaut dans le séquenceur Nuendo du même éditeur.

IR-1 : de Waves Ltd est un plug-in VST, TDM, RTAS ou Audio Unit pour Windows et Mac OS. Il propose une convolution « true stereo » et des réglages qui s'apparentent à ceux d'un effet de réverbération traditionnel qui viennent agir sur la réponse impulsionnelle.

Altiverb : de Audio Ease est sans doute la plus connue des réverbérations à convolution (VST, TDM, RTAS, Audio Unit...)

TL Space : de Avid est une réverbération à convolution dédiée uniquement à Protools (RTAS, TDM et audio suite)

Revolver : de Mc Dowell Signal Processing est une réverbération à convolution dédiée uniquement à Protools également, qui prend en charge les formats : AAX DSP, AAX Native, TDM, RTAS, AudioSuite et Audio Unit

F.3 Autres applications

Freeverb 3 : est une librairie open source de traitements audio qui contient des applications en lignes de commande pour Linux et des plug-ins VST pour Windows ou Mac OS dont Impulser 2 qui est une réverbération à convolution.

12. On parle d'un convolueur audio avec un fonctionnement « true stereo », lorsque chacun des deux canaux entrant passe par une convolution pour la voie de sortie gauche et pour la voie de sortie droite, ce qui fait quatre convolutions mono. (deux pour la gauche et deux pour la droite)

Convolver : est une application open source pour Windows pour l'application de filtres à réponse impulsionnelle finie (FIR) à de l'audio multicanal sous la forme d'un plug-in DSP temps réel pour Windows Media Player, de deux versions DirectX/DirectShow, d'un plug-in VST et d'applications en ligne de commande pour du traitement non temps réel.

Pure Data et Max/MSP : deux logiciels de programmation graphique orientés vers la création musicale et l'interaction temps réel. Ces logiciels permettent de programmer à peu près n'importe quelle fonction et intègrent des briques logiciels comme la convolution ou la FFT en natif ou bien par l'ajout de librairies externes supplémentaires. Pure Data est gratuit et open source tandis que Max/MSP, plus convivial est commercialisé par Cycling 74.

Web Audio API : une interface de programmation JavaScript de haut niveau pour le traitement et la génération de son dans des application web en HTML5 qui intègre un moteur de convolution. Ainsi, les applications JavaScript peuvent intégrer des traitements de convolution. Chrome le navigateur de Google est le premier à l'intégrer et devrait être rejoint par ses concurrents.



École Nationale Supérieure Louis Lumière 2013